National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

University of Alberta

The Structural Basis of DNA-Protein Recognition: A Nuclear Magnetic Resonance Determination of the Structure of DNA and DNA-Cro Repressor Interactions

by

James D. Baleja

A Thesis

Submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Biochemistry

Edmonton, Alberta

Spring, 1990

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60236-8

NL-339 (r. 88/04) c

Canada

University of Alberta
Edmonton
Canada T6G 2H7

Department of Biochemistry
474 Medical Sciences Building, Telephone (403) 492 3006

FEB 01 1990

Copyright Administrator
Journal of Magnetic Resonance
Academic Press
1 East First Street
Duluth, Minnesota 55802
U. S. A.

Dear Copyright Administrator,

A version of the *J. Magn. Reson.* manuscript, #23-177, entitled "Distance Measurement & Structure Refinement with NOE data" will form a chapter in my Ph. D. thesis on the structural basis of protein:DNA recognition using NMR techniques. The National Library of Canada will be preparing a microfilm copy of the thesis. I therefore request a letter of permission, for reproduction of copyrighted material.

Yours sincerely,

Jim Baleja

James D. Baleja

PLEASE TURN OVER

University of Alberta
Edmonton

Department of Biochemistry

Canada T6G 2H7

474 Medical Sciences Building, Telephone (403) 492-3006

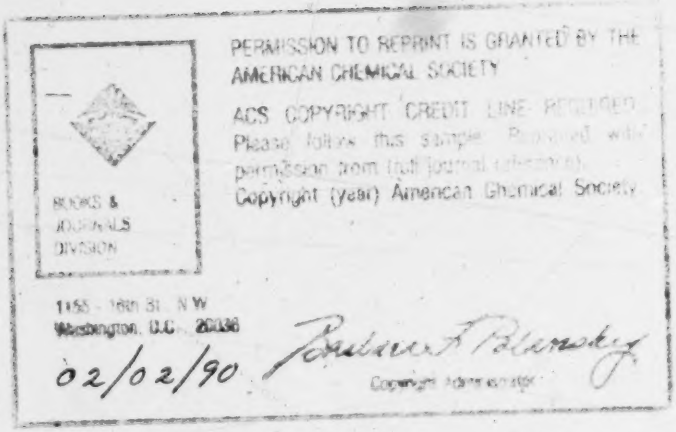
Copyright Administrator
Books and Journals Division
American Chemical Society
1155 Sixteenth Street, N. W.
Washington, D. C. 20036
U. S. A.

Dear Copyright Administrator,

A version of the *Biochemistry* manuscript, # BI 89 1963C, entitled "The Solution structure of Phage λ Half-operator DNA using NMR, Restrained Molecular Dynamics and NOE-based Refinement" will form a chapter in my Ph. D. thesis on the study of protein:DNA recognition using NMR techniques. The National Library of Canada will be preparing a microfilm copy of the thesis. I therefore request a letter of permission, for reproduction of the copyrighted material.

Yours sincerely,

James D. Baleja


PERMISSION TO REPRINT IS GRANTED BY THE AMERICAN CHEMICAL SOCIETY
ACS COPYRIGHT CREDIT LINE REQUIRED
Please follow this sample. Reprinted with permission from (full journal citation).
Copyright (year) American Chemical Society
BOOKS & JOURNALS DIVISION
1155 - 16th St. N.W.
Washington, D.C. 20036
02/02/90
Copyright Administrator

University of Alberta

Release Form

Name of Author: James D. Baleja

Title of Thesis: The Structural Basis of DNA-Protein Recognition: A Nuclear Magnetic Resonance Determination of the Structure of DNA and DNA-Cro Repressor Interactions

Degree: Doctor of Philosophy

Year this degree granted: 1990

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Jim Baleja

(Student's Signature)

c/o R. Baleja

(Student's Permanent Address)

P.O. Box 186

Sperling, MB

R0G 2M0

Date: March 27, 1990

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify they have read, and recommend to the Faculty of Graduate Studies and Reseach for acceptance, a thesis entitled The Structural Basis of DNA-protein Recognition: A Nuclear Magnetic Resonance Determination of the Structure of DNA and DNA-Cro Repressor Interactions

Submitted by James D. Baleja

In partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dr. Brian D. Sykes

Dr. Cyril M. Kay

Dr. A. Richard Morgan

Dr. J. William Lown

Dr. Wayne F. Anderson

Dr. Ponzy Lu

Date: March 7, 1990

Abstract

The structural basis of protein-DNA recognition is examined with an emphasis on the DNA-protein interactions of the Cro repressor protein from bacteriophage λ. The groundwork for the determination of nucleic acid structure in solution using nuclear magnetic resonance techniques is illustrated with mathematical procedures for distance determination, NOE data assessment, and NOE-based refinement. These NMR methods are used to determine the solution conformation of two self-complementary DNA octamers. The DNA duplexes are of alternating purine-pyrimidine sequence and demonstrate the diversity of DNA structure. Also determined is a DNA decamer, which comprises the left ten base-pairs of the O_R3 operator sequence from phage λ. All DNA structures are generally B type in character, but display local sequence-dependent variations in their conformation. The structure of the left ten base-pairs of O_R3 is compared to the corresponding DNA sequence from the crystallographic determination of the related λ cI repressor N-terminal domain-operator DNA complex. Despite the structure of the DNA being determined by different techniques, the conformation of the DNA is nearly the same, especially in regions not contacted by the protein. The interaction of this half-operator with wild-type and mutant Cro repressor proteins is detailed. A loss in DNA-binding ability of cross-linked Cro occurs in conjunction with a loss in flexibility. Formation of the complex between left ten base-pair half-operator and the native Cro repressor is studied. The O_R3 operator sequence is pseudo-symmetric, with its symmetry axis co-existing with the two-fold symmetry axis of the Cro protein. By characterizing the binding strengths and stoichiometries of different half-operator DNA sequences, the role of symmetry in protein-DNA interaction is highlighted. Implications of these investigations into the DNA-Cro repressor system from bacteriophage λ are discussed in terms of general protein-DNA recognition and of the techniques required for the study of protein-nucleic acid interaction.

Acknowledgments

I would like to acknowledge some of the many individuals who have helped with the work presented in this thesis. My supervisor, Brian Sykes, has given me the opportunity to carry out research in his laboratory and, where appropriate, has made invaluable criticisms to help improve the quality of the studies undertaken.

I also thank all in the lab for their support over the last few years. Dave Corson and Gerry McQuaid have maintained the spectrometers and magnets, and Linda Golden and Elke Metke have kept lab supplies in order. Helpful scientific discussions and/or restful mountain excursions were ongoing with Michelle Bjornsen, Joe O'Neil, Patricia Campbell, Brian Marsden, Heather Dettman, Gary Shaw, Frank Sönnichsen, Fred Brauer, and Dave Wishart. I am indebted to Wendy Findlay for her careful reading of manuscripts and especially to Gillian Henry for sound advice, patient attention, and for suggesting an occasional beer.

Others within the department have also helped me. In this regard, I thank Jenny van Eyk, Thor Borgford, Clive Sanders, Ruthven Lewis, Dave Heeley, Dave Mannock and Colin McCauley. I am appreciative for the advice given by Natalie Strynadka, Osnat Herzberg, Anita Sielecki and Mike James on structure determination.

I am grateful to Dr. Wayne Anderson for initiating the collaboration on the Cro project and for many helpful discussions on protein-nucleic acid interaction. I also thank J. Milne, G. Ozimek, and Carolyn Slupsky for their suggestions on Cro protein expression and purification.

Computer programs have been extensively used throughout this thesis. The GROMOS molecular simulation programs were supplied by Dr. W. van Gunsteren and Dr. Berendsen. Dr. R. Dickerson gave the helix analysis programs. Many of the graphics and DNA building programs have been written by D. Bacon. I very much appreciate the help of Robert Boyko for improvements in the NOE-based refinement subroutine and his many user-friendly data management utility programs. John Moult initiated me (and convinced me of the usefulness) in the use of computers for biochemical research.

I appreciate the materials and services obtained for these studies. DNA duplexes were supplied by W. Anderson, R. Pon, M. Germann, and J. H. van de Sande. Crude DNA duplexes were prepared by T. Atkinson. Plasmids encoding the native and mutant Cro repressors were obtained from the laboratory of Dr. M. H. Caruthers. Circular Dichroism measurements were made by K. Oikawa in the laboratory of Dr. C. Kay.

Coordinates for the X-ray determined structure of the λ repressor-DNA complex were obtained from Dr. C. Pabo. Dr. Alastair Muir has supplied me with the coordinates for the

Cro-DNA model and has helped generally in discussion of macromolecule structure and function.

Lastly, I would like to thank my parents, Roy and Vera, for their love, advice, and continual support.

Table of Contents

Chapter		Page
I.	Introduction	
	The Structural Basis of DNA-protein recognition	1
	A. DNA-protein interaction in biology	2
	B. Structure determination using nuclear magnetic resonance techniques	4
	C. DNA-protein recognition systems	
	1. Regulation of the mode of growth in bacteriophage λ	8
	a. Lysogeny	10
	b. A genetic switch	11
	c. The interaction of Cro repressor with DNA	12
	i. A model for the interaction of Cro with DNA	12
	ii. Testing the proposed Cro-DNA model	18
	d. The λ repressor from bacteriophage λ	28
	2. DNA recognition by the Cro and λ repressors of phage 434	32
	3. Protein-DNA Interactions in the Trp repressor-operator complex	35
	4. Studies of the Lac repressor and Lac repressor-DNA interactions	36
	5. The DNA-EcoRI endonuclease recognition complex	38
	6. Zinc finger DNA-binding proteins	41
	D. The structural basis of DNA-protein recognition	43
	1. General features of DNA-protein recognition	43
	2. Prospects of altering DNA-protein recognition	44
	3. Outline of the thesis	47
	E. References	47
II.	Nucleic Acid Structure Determination using NMR	51
	A. Introduction	52
	B. Theory	53
	C. Distance determination procedures	55
	1. Methods of distance determination using calculated NOE intensities.	55
	2. Distance and initial structure determination using observed NOEs	59
	D. NOE-based refinement	62

E. References..... 70

III. The Solution Conformation of Alternating Purine-Pyrimidine DNA using NMR, Restrained Molecular Dynamics, and NOE-based Refinement..... 72

A. Introduction..... 73

B. Experimental Procedures

1. Sample preparation..... 75
2. NMR Spectroscopy..... 77
3. Restrained Molecular Dynamics calculations..... 78
4. NOE-based structure refinement..... 79

C. Results

1. Resonance Assignment..... 81
2. Distance determination..... 91
3. Glycosidic dihedral angles..... 92
4. Right-handed DNA helix restraints..... 92
5. Molecular dynamics calculations and NOE-based refinement..... 93

D. Discussion

1. Structural details of alternating purine-pyrimidine DNA octamers.... 98
2. Biological role of alternating purine-pyrimidine sequences..... 104
3. Implication for structure determination of macromolecules in solution using NMR..... 105

E. References..... 107

IV. The Solution Conformation of Phage λ O_R3 Half-operator using NMR, Restrained Molecular Dynamics, and NOE-based Refinement..... 111

A. Introduction..... 112

B. Experimental Procedures

1. Sample preparation..... 114
2. NMR Spectroscopy..... 116
3. Restrained Molecular Dynamics 117
4. Structure refinement..... 117

C. Results

1. Resonance Assignment..... 119
2. Distance determination..... 119
3. Glycosidic dihedral angles..... 127
4. Right-handed DNA helix restraints..... 128

5. Molecular dynamics calculations 129
6. Structure refinement and assessment..... 133
7. Structural features..... 134
D. Discussion
1. Accuracy of structure determination..... 142
2. Extension to the full operator structure..... 144
3. Comparion between the conformations of L10 DNA when unbound and in complex with the λ repressor N-terminal domain..... 146
E. References..... 151

V. The Interaction of Phage λ O_R3 DNA with Native and V55C cro Repressor Proteins..... 154
A. Introduction..... 155
B. Experimental procedures
1. Materials..... 160
2. CD Spectroscopy..... 162
3. NMR Spectroscopy..... 162
4. Binding assays..... 163
C. Comparison of wild-type and V55C Cro proteins
1. Similarities in conformation and dynamics..... 165
2. Differences in structure and dynamics
a. Differences in thermal stability..... 176
b. Differences in amide exchange rates..... 179
3. DNA-binding characteristics
a. Interaction with O_R3 operator DNA..... 179
b. Interaction with O_R3 left half-operator..... 181
D. Formation and structural details of the Cro-L10 DNA complex..... 187
E. References..... 200

VI. Asymmetry in the Structure and Function of Phage λ Operator DNA..... 205
A. Introduction..... 206
B. Experimental procedures..... 208
C. Results..... 208
D. Discussion
1. The asymmetry of O_R3 operator DNA..... 213
2. Model for Cro-DNA aggregation..... 216

3. The interaction of Cro with O_R3 operator DNA..... 219
E. References..... 220

VII. Discussion
The Structural Basis of Operator DNA-Cro Repressor recognition..... 222
A. Summary..... 223
B. Future studies of protein:DNA recognition using NMR Spectroscopy
1. Amide exchange studies of Cro repressor..... 226
2. ^{31}P NMR of DNA..... 227
3. The interaction of Cro with other half-operators..... 227
4. RNA polymerase tail peptide..... 229
C. References..... 230

Appendix 1
FORTRAN subroutine for NOE restraint minimization..... 231

Appendix 2
Interproton distances in right-handed helices..... 244

Appendix 3
Experimental measurements and coordinates for $[d(GTACGTAC)]_2$..... 247

Appendix 4
Experimental measurements and coordinates for $[d(CATGCATG)]_2$..... 254

Appendix 5
Experimental measurements and coordinates for O_R3 half-operator,
d(TCTATCACCG)•d(CGGTGATAGTA), L10..... 261

List of Tables

Table I-1	Binding affinities of Cro proteins for DNA operators	46
Table II-1	Initial NOE build-up rates	56
Table II-2	Distance estimates	58
Table II-3	R factors for observed NOE intensities of L10	63
Table III-1	Proton chemical shift assignments for GTAC and CATG	90
Table III-2	Numbers and types of experimental restraints	91
Table III-3	Atomic rms differences between purine-pyrimidine structures	94
Table III-4	Energy terms for alternating purine-pyrimidine structures	97
Table IV-1	Proton chemical shift assignments for L10 DNA	126
Table IV-2	Atomic rms differences between L10 structures	130
Table IV-3	Potential energies (kJ/mol) for L10 structures	132
Table IV-4	Dihedral angles for L10	139
Table VI-1	^{1}H chemical shift assignments in Cro:DNA complexes	211
Table VII-1	Amide proton exchange rates in Cro repressors	227

List of Figures

Figure I-1	Protein synthesis	3
Figure I-2	^{1}H NMR spectrum of a protein	5
Figure I-3	Genetic map of λ	9
Figure I-4	Amino acid sequence of Cro repressor	12
Figure I-5	Regulatory operator DNA sequences of phage λ	13
Figure I-6	Base-pairs of DNA	15
Figure I-7	The O_R3 operator in a B DNA conformation	16
Figure I-8	Model of Cro-DNA interaction	17
Figure I-9	Sequence specific interactions between DNA and Cro repressor	19
Figure I-10	Amino acids homologies with Cro repressor	20
Figure I-11	A monomer of Cro repressor	22
Figure I-12	Effect of base substitutions in the binding of Cro to operator DNA	26
Figure I-13	Prediction of the binding affinity of Cro for DNA	27
Figure I-14	Structure of the λ repressor-DNA complex	30
Figure I-15	Sequence-specific interactions between DNA and λ repressor	31
Figure I-16	Lac represor operator	37
Figure I-17	Interactions between DNA and EcoRI endonuclease	39
Figure I-18	Zinc Finger amino acid sequence	42
Figure II-1	Evaluation of distance determination methods	57
Figure II-2	Distance estimate comparison	60
Figure II-3	Distance determination from observed NOE intensities	61
Figure II-4	Stereoview of the refined structure of L10 DNA	64
Figure II-5	Estimation of experimental error in NOE intensities	66
Figure II-6	Detection of motion by NOE measurements	67
Figure III-1	400 MHz ^{1}H NMR spectra of GTAC and CATG	76
Figure III-2	400 MHz ^{1}H NMR NOESY spectrum of GTAC	82
Figure III-3	Aromatic base<->sugar 1' region of NOESY spectra of GTAC and CATG	83
Figure III-4	Resonance assignment in nucleic acids	85
Figure III-5	Assignment of 2', 2", and 3' protons in 400 MHz COSY spectra of DNA octamers	86
Figure III-6	Base H6/H8 NOE connectivities to 3', 4', 5' and 5" protons	88

Figure III-7 NOESY and COSY spectra illustrating 3'<->4',5' and 5" connectivities in GTAC..... 89
Figure III-8 Structures of alternating purine-pyrimidine DNA..... 95
Figure III-9 Conformational parameters for alternating purine-pyrimidine DNA.... 99
Figure III-10 Base-stacking arrangements in GTAC..... 101
Figure III-11 Base-stacking arrangements in CATG..... 102
Figure III-12 Variations in the backbone and glycosidic bond torsion angles..... 103

Figure IV-1 One-dimensional 500 MHz ^1H NMR spectrum of L10 DNA..... 115
Figure IV-2 Assignment of base and 1' protons in L10..... 120
Figure IV-3 Assignment of 2', 2", and 3' protons in L10..... 121
Figure IV-4 Assignment of 3', 4', 5', and 5" protons in L10..... 122
Figure IV-5 Aromatic base<-> sugar 2',2" region of the NOESY spectrum..... 123
Figure IV-6 3'<->4' region of DQF-COSY spectrum at 30°C..... 124
Figure IV-7 One-dimensional NOE experiments in H_2O..... 125
Figure IV-8 Structural models for L10 DNA..... 131
Figure IV-9 Improvement in the NOE R factor..... 135
Figure IV-10 Variation in base roll, global helix twist, propellor twist, and δ_{1-2} angles for the refined structures of L10 DNA..... 136
Figure IV-11 Variation in rise, displacement, slide, and tilt angles for the refined structures of L10 DNA..... 137
Figure IV-12 Base-stacking arrangements in L10 DNA..... 141
Figure IV-13 Expansion of 2'<->3' DQF-COSY crosspeaks for residues C_2 and T_3 of L10 DNA..... 143
Figure IV-14 Models for the 21 base-pair O_R3 operator DNA..... 145
Figure IV-15 Structure of DNA free and in complex with λ repressor..... 147
Figure IV-16 Atomic rms deviation between free and bound operator DNA..... 149
Figure IV-17 Comparison of conformation in free and complexed DNA..... 150

Figure V-1 O_R3 and L10 DNA sequences..... 156
Figure V-2 The valine to cysteine substitution in Cro repressor..... 159
Figure V-3 Comparison of ^1H NMR spectra for native and V55C Cro protein..... 166
Figure V-4 ^1H NMR spectra of native and V55C Cro aromatic side-chains..... 167
Figure V-5 ^1H NMR spectra of native and V55C Cro α protons..... 168
Figure V-6 ^1H NMR spectra of native and V55C Cro methyl protons..... 169
Figure V-7 Antiparallel β sheet structure..... 171

Figure V-8 NOESY spectra of native and V55C Cro proteins..... 172
Figure V-9 DQF-COSY spectrum of V55C Cro protein..... 173
Figure V-10 Structure of the β sheet core of V55C Cro protein..... 174
Figure V-11 CD spectra of native and V55C Cro repressor proteins..... 175
Figure V-12 Observation of thermal denaturation by NMR..... 177
Figure V-13 Thermal stabilities of native and V55C Cro repressor proteins..... 178
Figure V-14 CD spectra Cro-operator DNA complexes..... 180
Figure V-15 Imino proton NMR resonances of free and complexed DNA..... 182
Figure V-16 Titration curves of L10 half-operator by Cro proteins..... 183
Figure V-17 Magnitude of induced imino chemical shifts upon complex formation. 185
Figure V-18 Comparison of wild-type and V55C Cro proteins in protein:DNA complexes..... 186
Figure V-19 ^{1}H NMR spectra of Cro, L10 DNA, and the protein:DNA complex.... 188
Figure V-20 Spectra changes in L10 DNA H6/H8 resonances upon addition of Cro protein..... 189
Figure V-21 Titration of L10 DNA by Cro repressor..... 191
Figure V-22 Linewidth changes in DNA upon addition of Cro repressor..... 193
Figure V-23 Estimation of differences in chemical shift..... 194
Figure V-24 DNA aromatic base chemical shift changes upon addition of Cro repressor..... 195
Figure V-25 NOESY spectrum of unbound Cro repressor..... 197
Figure V-26 NOESY spectrum of the Cro:DNA complex..... 198

Figure VI-1 DNA half-operator sequences..... 207
Figure VI-2 Titration of Cro aromatic side-chain ^{1}H NMR resonances with DNA.. 209
Figure VI-3 Titration shifts of Cro ^{1}H NMR resonances with R9 DNA..... 210
Figure VI-4 Titration of Cro methyl ^{1}H NMR resonances with DNA..... 216
Figure VI-5 Model for the aggregation of 1:1 L10 DNA:Cro protein complexes.... 218

Figure VII-1 ^{31}P NMR spectrum of L10 DNA..... 228

List of Abbreviations and Symbols

CATG	$[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ DNA
CD	circular dichroism
CIDNP	Chemically Induced Nuclear Polarization
COSY	Two-dimensional correlated spectroscopy
DQF-COSY	Double-quantum filtered COSY
DSS	disodium-2,2-dimethyl(-2-silapentane-5-sulphonate)
EDTA	ethylenediaminietetraacetic acid
GTAC	$[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ DNA
IPTG	isopropylthiogalactoside
J	NMR spin-spin coupling constant
L10	d(T-C-T-A-T-C-A-C-C-G)•d(C-G-G-T-G-A-T-A-G-A) DNA
MD	molecular dynamics
NMR	Nuclear Magnetic Resonance Spectroscopy
NOE	Nuclear Overhauser Effect or Enhancement
NOEs	NOE intensities
NOESY	Two-dimensional NOE spectroscopy
OD_{600}	Optical density at 600 nanometers
rms	root mean square
SDS	sodium dodecyl sulfate
TRIS	tris(hydroxymethyl)aminomethane
δ	NMR Chemical shift

dA	Adenine, deoxyadenosine
dC	Cytosine, deoxycytidine
dG	Gaunine, deoxygaunosine
dT	Thymine, deoxythymidine

A, Ala	Alanine
C, Cys	Cysteine
D, Asp	Aspartate
E, Glu	Glutamate
F, Phe	Phenylalanine
G, Gly	Glycine
H, His	Histidine

I, Ile	Isoleucine
K, Lys	Lysine
L, Leu	Leucine
M, Met	Methionine
N, Asn	Asparagine
P, Pro	Proline
Q, Gln	Glutamine
R, Arg	Arginine
S, Ser	Serine
T, Thr	Threonine
V, Val	Valine
W, Typ	Tryptophan

Chapter I

The Structural Basis of DNA-protein ' ition

A. Introduction

DNA-protein recognition plays a central role in biology. The association of protein with DNA has been shown to regulate cellular events as diverse as metabolite utilization, growth and development, and cell differentiation. A better understanding of the way in which proteins modify transcriptional activity will lead to *de novo* methods of correcting aberrant processes in human disease, and of altering the genetic expression of commercially applicable biological processes to effect designed and beneficial changes for society.

Proteins are important in nearly all biological processes—in catalysis, transport, coordinated motion, excitability, and the control of growth and differentiation. Their level of activity can modulated in two ways—by regulating the ability of the protein to carry out its function, for example, by adding co-factors or by product inhibition, and by regulating the quantity of protein produced. The information encoding the kinds and amounts of proteins to be made is contained in the DNA of the cell. The genetic information encoding the protein is first transcribed by RNA polymerase to produce an intermediate RNA molecule (Figure I-1). The binding site of RNA polymerase on DNA is termed the promoter. Synthesis of proteins according to the information content of the RNA molecule is called translation, and is carried out by ribosomes. Certain proteins can bind at specific locations on the DNA, called operator sites, and regulate the transcriptional activity at or near the site of interaction. It is the purpose of this thesis to explain how protein and DNA recognize each other.

In this introduction, the different mechanisms by which proteins and DNA recognize each other and the diverse techniques used to study protein-DNA interaction are illustrated. Particular emphasis is placed on the use of nuclear magnetic resonance (NMR) techniques which is used in subsequent chapters to study the Cro repressor protein-O_R3 operator DNA system from the bacteriophage λ. Recent crystallographic determinations of DNA-protein complexes (involving the λ repressor, phage 434 repressor, phage 434 Cro, Trp repressor, and EcoRI endonuclease proteins) have contributed greatly to an understanding of protein-DNA recognition. Also reviewed are two systems for which NMR techniques have contributed significantly—the Zn finger proteins, and the *Escherichia coli* Lac repressor protein. In particular, the structural and systematic mutational analysis work from this decade (1980-1989) is emphasized, with much of the previous elegant biochemical and genetic studies either referred to or summarized briefly.

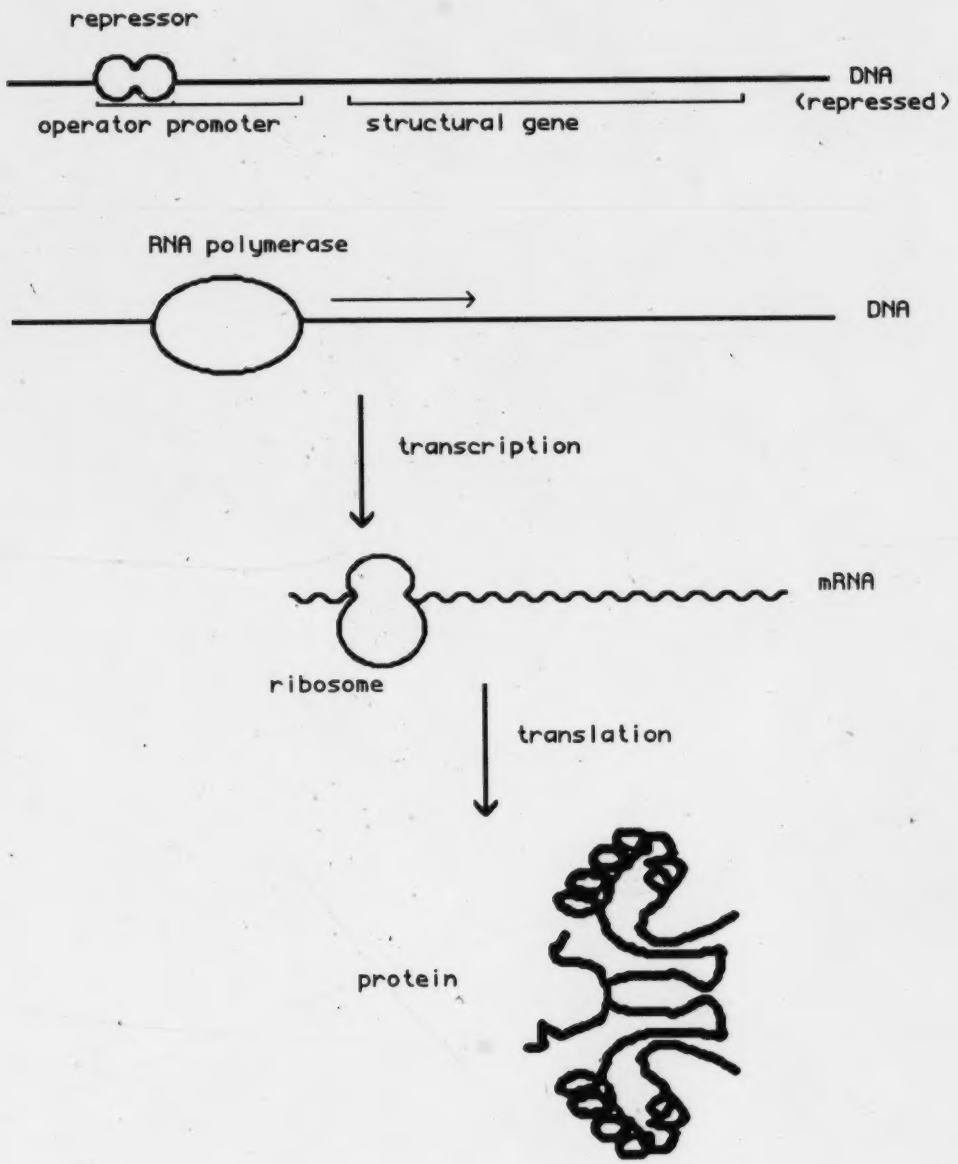


Figure I-1 Protein synthesis

Information encoding the amino acid sequence of a protein is stored in the DNA base sequence of a structural gene. RNA polymerase binds at the promoter sequence to transcribe a RNA copy of the DNA sequence. Transcription can be regulated by repressor proteins that bind to the DNA at operator sequences. Proteins are synthesized by ribosomes that read the information content in the RNA and translate it into the amino acid sequence.

B. Structure determination using nuclear magnetic resonance techniques

Dramatic improvements in NMR instrumentation and methodology have made it possible to determine the three-dimensional structures of small proteins and nucleic acids in solution. It has been gratifying to see a large part of this revolution in structure determination occur during my career as a Ph. D. student. My intention is not to discuss the mechanics of structure determination using NMR (this is done in a monograph by Wüthrich, 1986; is reviewed by Bax, 1989; van de Ven & Hilbers, 1988; and examples are given in subsequent chapters of this thesis) nor the modern NMR techniques used in this process (see Bax, 1982; Morris, 1986; Ernst et al., 1987). However, an outline of the procedure is given, and the most important recent developments are highlighted.

Structure determination using NMR generally consists of three sequential steps: (1), assignment of resonances in the ^1H NMR spectrum to specific nuclei of the molecule; (2), estimation of interproton distances by measuring the extent of the dipolar relaxation pathway between protons with the nuclear Overhauser effect; and (3), the incorporation of the spatial information to generate a model by a method such as distance geometry or restrained molecular dynamics.

A one-dimensional ^1H NMR spectrum of a protein (Cro repressor) is shown in Figure I-2. Each resonance originates from a particular proton of the protein. The resonance position in the spectrum reflects the shielding of the nucleus by surrounding electrons, with decreasing resonance frequency corresponding to more shielded protons (towards the right of the spectrum). Backbone amide and aromatic side-chain protons are the most de-shielded and resonate about 6 to 10 ppm higher relatively to a commonly used reference signal from the methyl protons in DSS (dimethylsilapentanesulfonate). Protons on the α carbon typically resonate between 3 and 5 ppm, methylene protons between 1 and 3.5 ppm, and methyl protons between 0 and 2 ppm. Because of the unique spatial fold of the protein and the different chemical nature of the twenty amino acid residue types, there are differences in chemical shift, for example, between two α protons. The assignment problem is to find the resonance frequency for each proton of the macromolecule.

Resonance assignment first requires the observation of non-overlapping resonances in the NMR spectrum. The most important methodological development has been the introduction of multi-dimensional NMR, spreading the severely overlapping one-dimensional NMR spectrum into two, and now, three, orthogonal frequency dimensions. Key instrumental improvements have been the introduction of computers to handle the experimental data, and higher-field instruments to further increase spectral resolution. The increase in field strength and improved radio-frequency technology has also led to a large

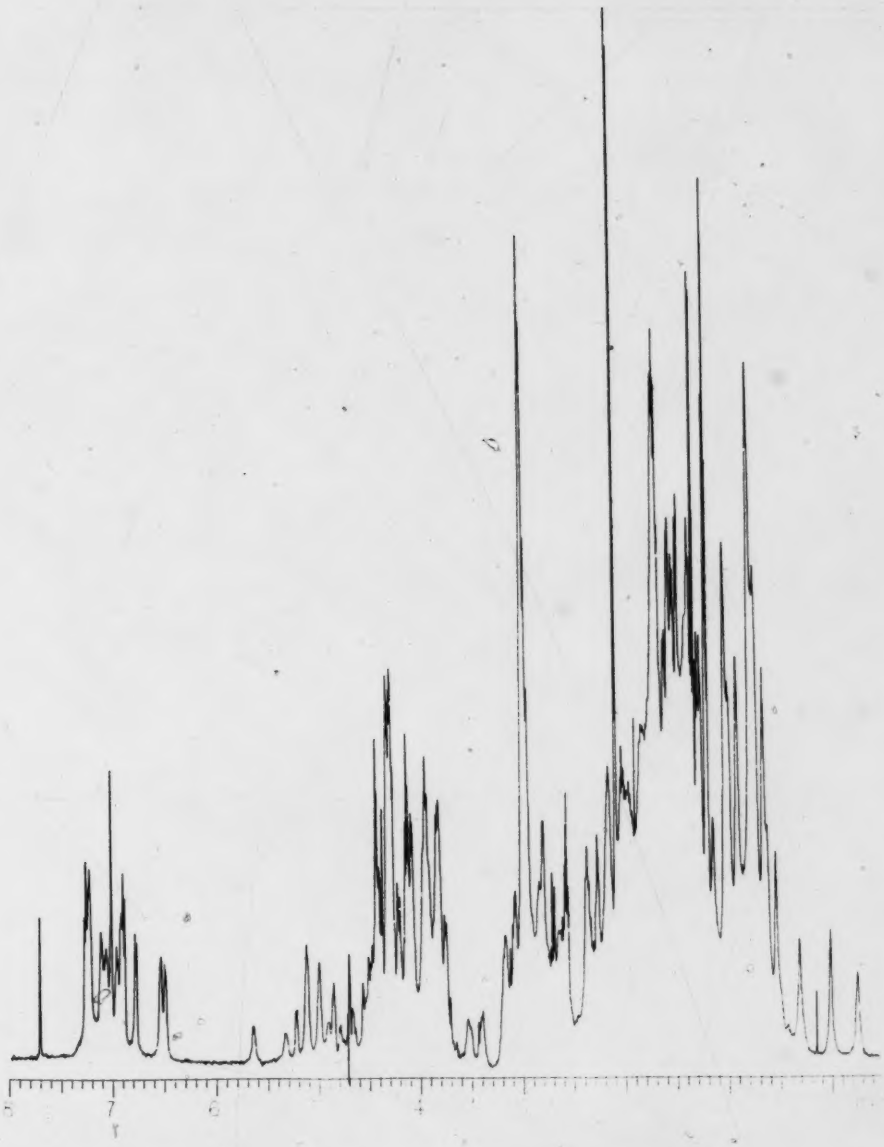


Figure I-2 One-dimensional ^1H NMR spectrum of Cro repressor protein.

The one-dimensional spectrum of the Cro repressor protein was measured at 500 MHz on a sample concentration of 0.75 mM in a buffer of 50 mM KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, pH 6.9 in D_2O at 30°C. Residual HOD was suppressed by decoupling during the relaxation delay (1.8 sec.) between scans (16 in total). Backbone amide protons have been exchanged for solvent deuterons (and normally resonate between 6 and 10 ppm). Resonances from aromatic side-chain protons are between 6 and 8 ppm. Aliphatic protons resonate between 0 and 6 ppm (see text for details).

increase in NMR sensitivity, enabling structural studies at millimolar concentrations (and less). A sample concentration of several millimolar is required on our 1982 'vintage' Nicolet 300 MHz spectrometer for the acquisition of two-dimensional spectra with sufficient quality for structure determination. By way of contrast, the structure determination of DNA duplexes was carried out on a Varian XL-400 spectrometer (*circa* 1985) at submillimolar concentration (Chapter III). Further improvements in sensitivity and spectral dispersion with the installation of a Varian VXR-500 spectrometer in 1986 permitted studies of larger DNA duplexes (Chapter IV) and of their interaction with DNA-binding proteins (Chapters V and VI). Procedures that exploit the correlation of resonance frequencies through the spectral properties of nuclei involved have been developed and have become the method of choice for resonance assignment (reviewed by Wüthrich, 1986; Bax, 1989).

The interaction of nuclei in a molecule allows the design of NMR experiments to correlate resonance frequencies with each other. Different information can be obtained for the conformational relationship of two protons, depending on the type of correlation measured. For example, two protons that are separated by three bonds have a spin-spin coupling constant which can be measured by resonance line splitting in one-dimensional spectra, and by the observation of cross-peaks in two-dimensional correlation spectroscopy (COSY) spectra. The coupling constant, J, has a characteristic dependence on the dihedral angle, θ:

$$J(\theta) = A\cos^2\theta - B\cos\theta + C \qquad (1)$$

where A, B, and C are empirically determined constants (reviewed by Bax, 1989). Measurement of coupling constants gives dihedral angle information in a molecule. More useful is the NOE experiment which measures the rate at which protons exchange their nuclear magnetization. Perturbation of the magnetization of a nuclear spin, i, affects the magnetization of a spatially proximate spin, j, in a time-dependent manner. At short times, the effect is inversely proportional to the sixth power of the distance between the spins:

$$NOE = K \cdot r_{ij}^{-6} \qquad (2)$$

where the constant K is dependent on the rotational correlation time between spins i and j. This constant is typically determined using the observed NOE for a proton pair of known distance, such as the 1.76 Å separation of two methylene protons. Because the dependence on the sixth power of the distance, the NOE is only observable for two protons less than 5 Å or so. Nonetheless, several hundred crosspeaks are often observed in two-dimensional

NOE (NOESY) spectra, and the intensities of the crosspeaks between resonance frequencies can be used to determine the distance between the two protons involved.

The determination of structure with high precision requires as many distances as possible, and these distances to be as accurate as possible. Because the NOE intensity can be measured with an accuracy of about 10%, one might expect distances to be determined with very high accuracy (±1.5%, Bax, 1989). In practice, local internal motions give rise to different rotational correlation times for different interproton vectors, and the NOE between two protons is often modified by indirect or relayed NOE effects. The effects of motion can be modelled by a differential motion model to take into account the presence of fast local motions (Chapter II; Baleja et al., 1990). With the availability of faster computers, indirect NOE contributions can be taken into account (also discussed in Chapter II).

However, the low ratio of the number of observables over the number of conformational parameters reflects one of the problems for structure determination using NMR techniques. The NOE-based refinement method (Chapter II) has the potential to use NMR parameters that cover a larger region of conformational space, and therefore to better determine structures in solution using NMR. Structure determination methods (reviewed by Kaptein, 1988) generally incorporate standard geometrical constraints, such as bond lengths, bond angles, and van der Waals radii, to generate structures consistent with the experimental data. Restrained molecular dynamics simulations are used in Chapter III and IV to generate nucleic acid structures.

Further advances in NMR structure determination technology will ensure its place as a method of choice for the determination of structure to high accuracy in solution. For suitable small proteins, the structure can sometimes be determined in less than a few months (Bax, 1989). Even at relatively low resolution, NMR structures offer valuable complementary information to those solved by X-ray crystallographic means, and have sometimes indicated crystal packing perturbations in the crystal structure. Larger proteins are being studied by the incorporation of NMR-active stable isotopes (^{15}N, ^{13}C, ^{2}H, ^{19}F) which result in spectral simplification and in easier resonance assignment. Since NMR can probe motions over a time scale spanning ten orders of magnitude, it can provide significant insights into the role of protein dynamics with respect to their function in biology. NMR enables researchers, for the first time, the opportunity to study the structures of biological macromolecules in solution. It has been a pleasure to contribute to and to observe the development of the field. Later in this chapter, I shall illustrate the results using NMR for structure determination for the Lac repressor headpiece and its complex with DNA.

C. DNA-protein Recognition Systems

1. The regulation of the mode of growth in bacteriophage λ.

To indicate the importance of protein-DNA recognition in biology, the regulatory system concerning the mode of development from bacteriophage λ is discussed in some detail. The λ phage has a double-stranded DNA genome organized into functional units (Friedman et al., 1984). As a temperate phage, it can adopt either one of two mutually exclusive life styles: lytic or lysogenic. Upon infection, phage DNA is injected into the host bacterium, *Escherichia coli*. In the lytic pathway, phage gene expression leads to the production of phage particles and the lysis of the bacterium. In the lysogenic pathway, two separate events occur—the phage genome recombines with the bacterial chromosone to integrate the phage DNA into the host chromosone, and the phage produces a repressor that binds to particular sites on the phage DNA to shut off synthesis of most phage-encoded products. The integrated phage, called a prophage, is maintained by the bacterium in a quiescent state until an inducing event such as DNA damage occurs. Upon induction, phage repression is removed, the prophage is excised from the bacterial chromosome, and phage particles are produced, as in a lytic infection.

The decision between lysis and lysogeny immediately following infection of the host bacterium is dependent on two main environmental factors: nutritional conditions of the host and the multiplicity of infection. The establishment of either the lytic or lysogenic mode involves the interaction with several host proteins, and has been extensively reviewed (Friedman et al., 1984; Hendrix et al., 1983). Here, the molecular events involving the maintenance of the lysogenic state and the induction of the lytic cycle are described to provide an insight into the significance of protein-DNA interaction in biology.

The bacteriophage λ encodes two proteins that play contrasting roles in determining its mode of growth. One of these is the product of the cI gene, and is classically called the λ repressor. It is required for lysogeny. The other is the product of the *cro* gene, herein called Cro repressor, and is required for lytic growth.

The effects of λ repressor and Cro repressor result from their differential binding to two operator regions on the phage chromosone (Figure I-3).

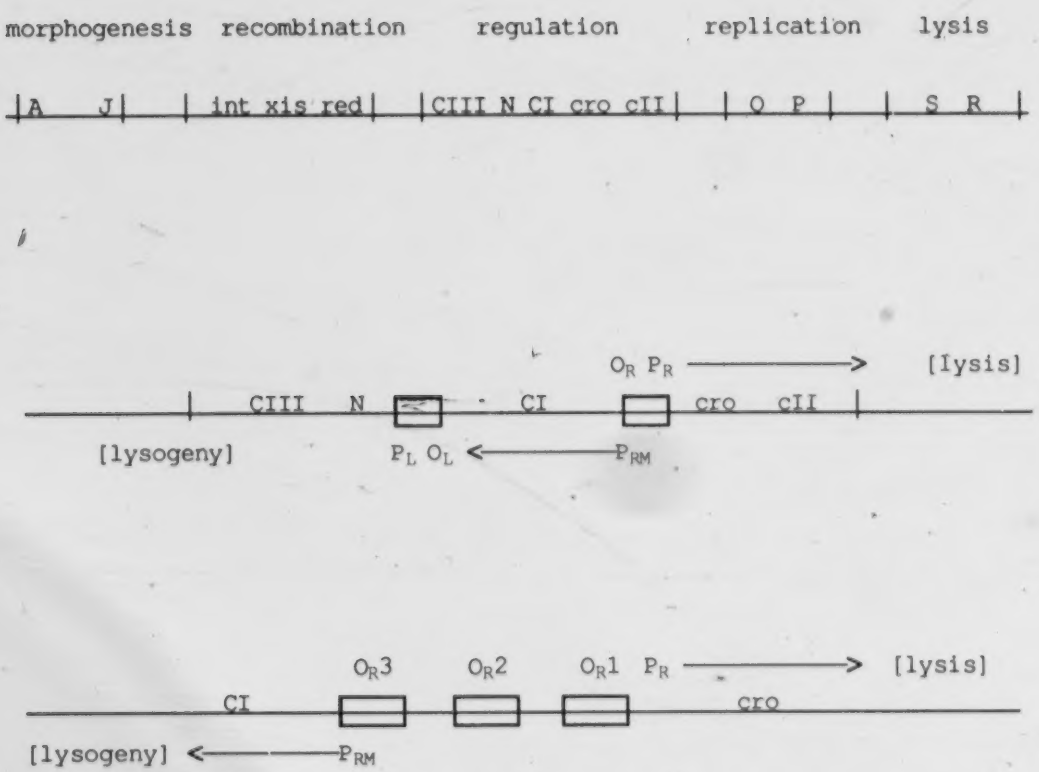


Figure I-3 Genetic map of phage λ

The top line of the figure outlines the genome of bacteriophage λ. Representative genes are given for each functional group (adapted from Friedman et al., 1984). The genes associated with regulation of the life cycle of λ are illustrated in the middle. The bottom line shows an expansion of the regulatory region around the right operator region. Cro repressor binding at O_R3 prevents transcription required for the lysogenic pathway. The product of the *cI* gene, called λ repressor, binds most strongly to the O_R1 operator site and prevents transcription of the *cro* gene.

The left operator, O_L, controls transcription from a set of early genes. The primary events regulating the maintenance of the lysogenic state and induction of the lytic state involves interactions of proteins with the right operator region O_R. Within this region there are three similar, but not identical DNA sequences, O_R1-3, which are the preferred binding sites for the λ and Cro repressors. A repressor protein binding at O_R3 prevents RNA polymerase from binding at P_{RM}, and therefore leftward transcription of the cI gene. A repressor protein binding at O_R1 overlaps with the P_R promoter, preventing initiation of rightward transcription of the *cro* gene. The differential binding affinities of λ and Cro repressor for these three sites explains their contrasting roles in the life cycle of λ. λ repressor binds most readily to the O_R1 and O_R2 sites, O_R3 only being filled at higher concentrations of protein. Cro repressor prefers O_R3, and only fills the remaining sites when Cro levels are high.

a. Lysogeny

Under conditions of lysogeny, there are approximately 200 molecules of λ repressor per cell, and nearly zero Cro repressor (Shea & Ackers, 1985). λ repressor binding at O_R1 prevents Cro synthesis and the lytic part of the life cycle. When λ repressor is bound at O_R2, transcription from P_{RM} is activated (Hochschild et al., 1983). Transcription is not enhanced by an induced change in DNA structure downstream from O_R2, but by presenting sites on the protein additional to the promoter site which increase RNA polymerase binding. If λ repressor levels become too high, the O_R3 site is filled, also repressing P_{RM}, and temporarily slowing the synthesis of λ repressor mRNA[1]. Therefore, the λ repressor acts both as a repressor of (mainly) *cro* transcription, and activates its own. The level of λ repressor present in a lysogen is sufficient to prevent expression of most of the viral DNA functions until activated by an inducing signal in the environment. This suppression also prevents other phage from super-infecting the bacterium, leading to a type of immunity.

---

[1]However, concentrations of λ repressor rarely reach this level *in vivo*, so that autogenous negative control is most likely not significant in stabilizing the repressor level in a lysogen. Low levels of cro (from transcription leakage) are more likely important.

b. A genetic switch—induction to the lytic state from the lysogenic state

The λ prophage is essentially an inert piece of genetic material kept dormant by virtue of the λ repressor. The central event in induction is the removal of repression, which is accomplished by destruction of the λ repressor. Now in the lytic state, the phage genome multiplies autonomously and phage components are synthesized and assemblied. The lytic events culminate in the production of infectious phage particles which are eventually released from the lysed host cell.

The best studied system of induction is by exposure of the bacterial lysogen to ultra-violet light, or various compounds known to alter DNA metabolism. *E. coli* reacts to DNA-damaging stimuli by a mechanism known as the SOS response. The single-stranded DNA formed as a result of DNA damage is a co-factor for a host protein, the *recA* gene product, which has several DNA recombination and repair properties. It also activates a proteolytic activity associated with RecA protein, which , in turn, results in activation of expression for a set of damage inducible genes, enabling the bacterium to repair damage. The SOS response also signals the possibility of the imminent demise of the host, and therefore the demise of the prophage. To save itself, the prophage must be able to heed a strong SOS signal (RecA protein), and escape the dying host by growing lytically.

At high levels of DNA damage, when host survival is clearly imperiled, the RecA protein stimulates cleavage of the λ repressor. Cleaved λ repressor has sufficiently reduced affinity for the O_R1 site so that RNA polymerase can now transcribe at P_R. A burst in the level of Cro repressor results. The Cro repressor binds at the O_R3 site, preventing any further production of λ repressor and this commits the cell to the lytic cycle of phage development. The dormant phage can now become virulent.

Regulation of the life cycle of phage λ is a clear example of one of many biochemical processes governed by protein-DNA interaction (Jacob & Monod, 1961). The mechanisms by which Cro repressor protein distinguishes O_R3 from other DNA sites are not only of interest as a focal point of this thesis, but also as a general model for protein-DNA interaction.

c. The interaction of Cro repressor with DNA

i. A model for the interaction of Cro with B DNA

Much of the work in deciphering how the Cro repressor recognizes DNA began with its purification (Folkmanis et al., 1976). Subsequent assessment of *in vitro* properties and amino acid sequence (Takeda et al., 1977; Hsiang et al., 1977; Roberts et al., 1977; Figure I-4) revealed a small basic dimeric protein of identical subunits of molecular weight 7351. Filter-binding assays show that it bound most tightly to the DNA containing the O_L and O_R operator sites, but that it also bound other DNA sequences, but with up to several orders of magnitude lower affinity. DNAase protection studies (reviewed by Ptashne et al., 1980), show six major binding sites, three each in the O_R and O_L operator regions. These DNA sequences are shown in Figure I-5. These sites are similar, but not identical, and Cro repressor affinity for the different sequences varies over a near hundred fold range. Accounting for differences in binding affinity between these six sites will provide a explanation for much of the lower affinity of Cro for non-operator DNA.

1				5					10
Met	Glu	Gln	Arg	Ile	Thr	Leu	Lys	Asp	Tyr
				15					20
Ala	Met	Arg	Phe	Gly	Gln	Thr	Lys	Thr	Ala
				25					30
Lys	Asp	Leu	Gly	Val	Tyr	Gln	Ser	Ala	Ile
				35					40
Asn	Lys	Ala	Ile	His	Ala	Gly	Arg	Lys	Ile
				45					50
Phe	Leu	Thr	Ile	Asn	Ala	Asp	Gly	Ser	Val
				55					60
Tyr	Ala	Glu	Glu	Val	Lys	Pro	Phe	Pro	Ser
				65	66				
Asn	Lys	Lys	Thr	Thr	Ala				

Figure I-4 The amino acid sequence of Cro repressor protein
The amino acid sequence of Cro has been determined by directly sequencing the protein (Hsiang et al., 1977) and has been deduced from the DNA sequence (Roberts et al., 1977). Basic residues are underlined.

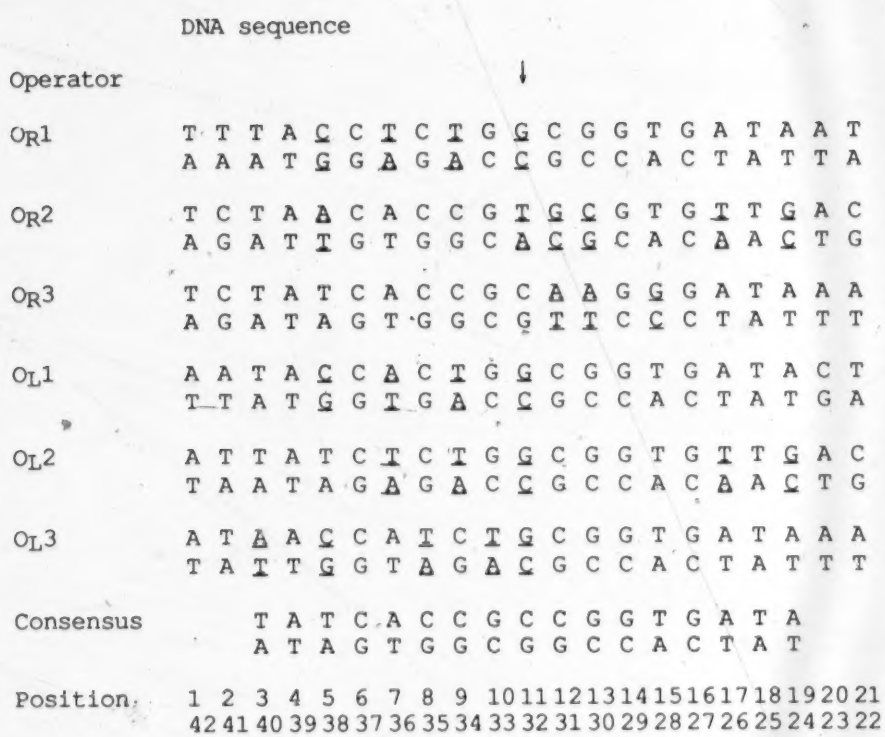

DNA sequence
Operator
O_R1 T T T A C C T C T G G C G G T G A T A A T
A A A T G G A G A C C G C C A C T A T T A
O_R2 T C T A A C A C C G T G C G T G T T G A C
A G A T T G T G G C A C G C A C A A C T G
O_R3 T C T A T C A C C G C A A G G G A T A A A
A G A T A G T G G C G T T C C C T A T T T
O_L1 A A T A C C A C T G G C G G T G A T A C T
T T A T G G T G A C C G C C A C T A T G A
O_L2 A T T A T C T C T G G C G G T G T T G A C
T A A T A G A G A C C G C C A C A A C T G
O_L3 A T A A C C A T C T G C G G T G A T A A A
T A T T G G T A G A C G C C A C T A T T T
Consensus T A T C A C C G C C G G T G A T A
A T A G T G G C G G C C A C T A T
Position 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22

Figure I-5 Regulatory operator DNA sequences of phage λ

Sequences of the O_R and O_L regulatory operators as they occur in the bacteriophage λ genome. The arrow at the top corresponds to the central base-pair at position 11-32. The sequences have approximate two-fold symmetry about this base-pair. The sequence of the two-fold symmetric consensus operator is derived from the most commonly observed base sequences in each half of operators in O_R and O_L. Bases different from the consensus sequence within the middle 17 base-pairs are underlined. Footprinting studies with Cro or λ repressor on whole λ genomic DNA show that the proteins protect only the middle 17 base-pairs of these operators. However, model building studies indicate possible interactions with phosphates over 19 or 20 bases. Including 21 base-pairs also avoids end effects in studies with oligonucleotides.

Dimethyl sulfate methylates the N7 of guanine and the N3 of adenine in double-stranded DNA. These are located in the major and minor grooves, respectively (Figure I-6). Cro protects guanines, but not adenines from methylation, indicating that the protein interacts with DNA in the major groove. In related experiments, ethylnitrosourea was used to probe which phosphate groups are important in Cro binding. In solution, DNA exists predominantly in a classical B DNA conformation (Arnott & Hukins, 1972). A B DNA model of the O_R3 operator is shown in Figure I-7. The pattern of the guanine and phosphate functional groups that interact with Cro are along one face of the DNA helix, suggesting that the protein binds along one side of the DNA, and does not encircle it. The sequence of each DNA site is approximately two-fold symmetric (Figure I-5). The dimeric nature of Cro and the chemical probe experiments therefore suggest that each domain or monomer of repressor contacts only (or predominantly) the left or right half of an operator site. Nearly the same interactions are expected to occur for both halves of an operator and Cro dimer.

Further details of the interaction of Cro and DNA became apparent with the determination of the three-dimensional structure of the Cro protein using crystallographic methods (Anderson et al., 1981). A pair of two-fold related α helices are seen to protrude from the protein molecule (Figure I-8). They are approximately 34 Å apart, or the same distance as two successive major grooves of DNA. Furthermore, the angle made between the axis of the two α-helices and a line connecting their mid-points is equal to the angle between the major grooves of B DNA and the overall DNA helix axis. A model was prepared by placing the Cro dimer so that its two fold axis of symmetry corresponded to that of the DNA, and adjusting the relative positioning of the protein and DNA. The DNA was, in addition, allowed to curve and the twist per base-pair (approximately 34.6° in B DNA) was permitted to vary in order to optimize the interaction between protein and DNA. In the model, the amino acid side-chains of the residues of the recognition helix (α-3) make specific hydrogen bonds and van der Waals contacts with the edges of base-pairs exposed in the major groove of B DNA. A number of sequence-independent interactions mainly to the phosphate backbone stabilize the Cro-DNA complex. These mainly involve residues of the second helix, the 3rd strand of the β sheet, and the flexible carboxy terminus.

This model was consistent with the chemical modification studies, and also explained the reduced affinities of Cro for known naturally occurring O_R3 mutants. Because the number of non-specific sites is much greater than the operator sites, Cro is most likely to first bind to an arbitrary site. When contained within large stretch of DNA, the association of operator DNA with Cro is a fast reaction—at least 10-100 times faster than the rate associated from simple diffusion (Kim et al., 1987). The increase in rate of association is

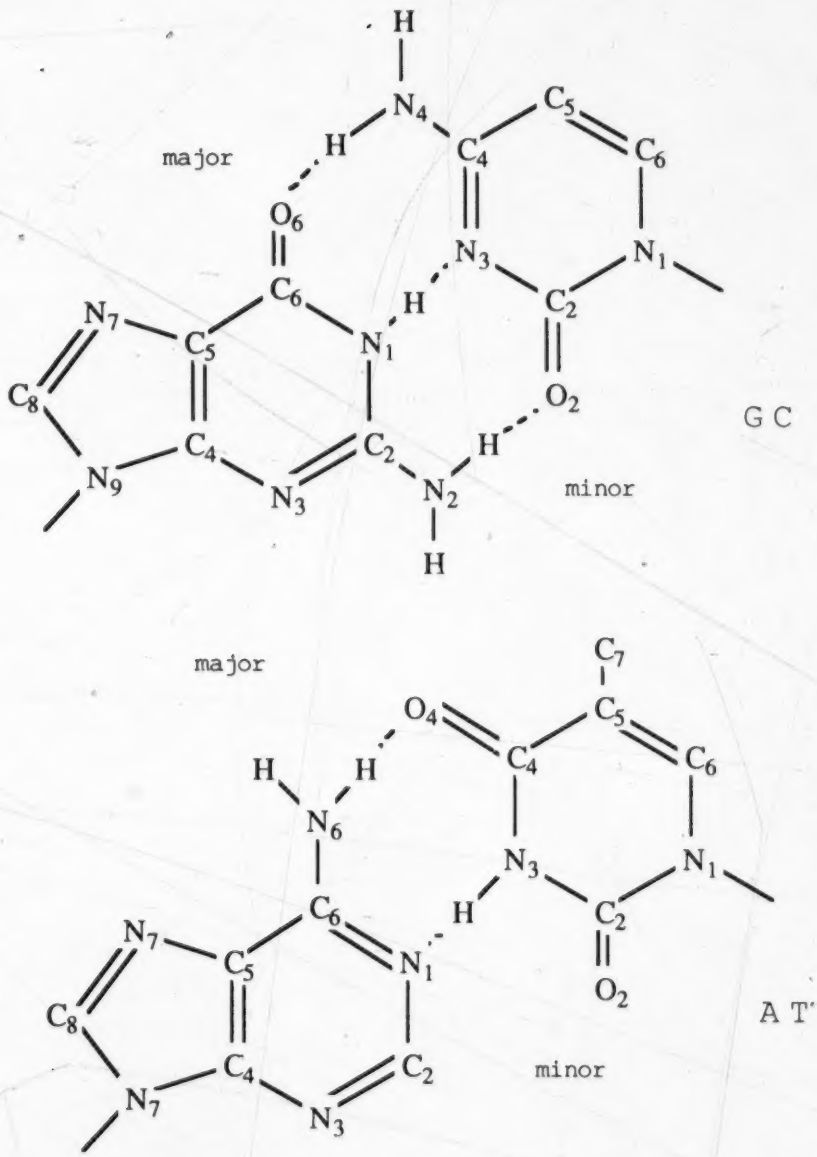


Figure I-6 Base-pairs of DNA

The bottom of each Watson-Crick base-pair forms a minor groove when in a helix. The top part forms the major groove. Dimethylsulfate can alkylate N7 of guanine, in the major groove, and N3 of adenine, in the minor groove.

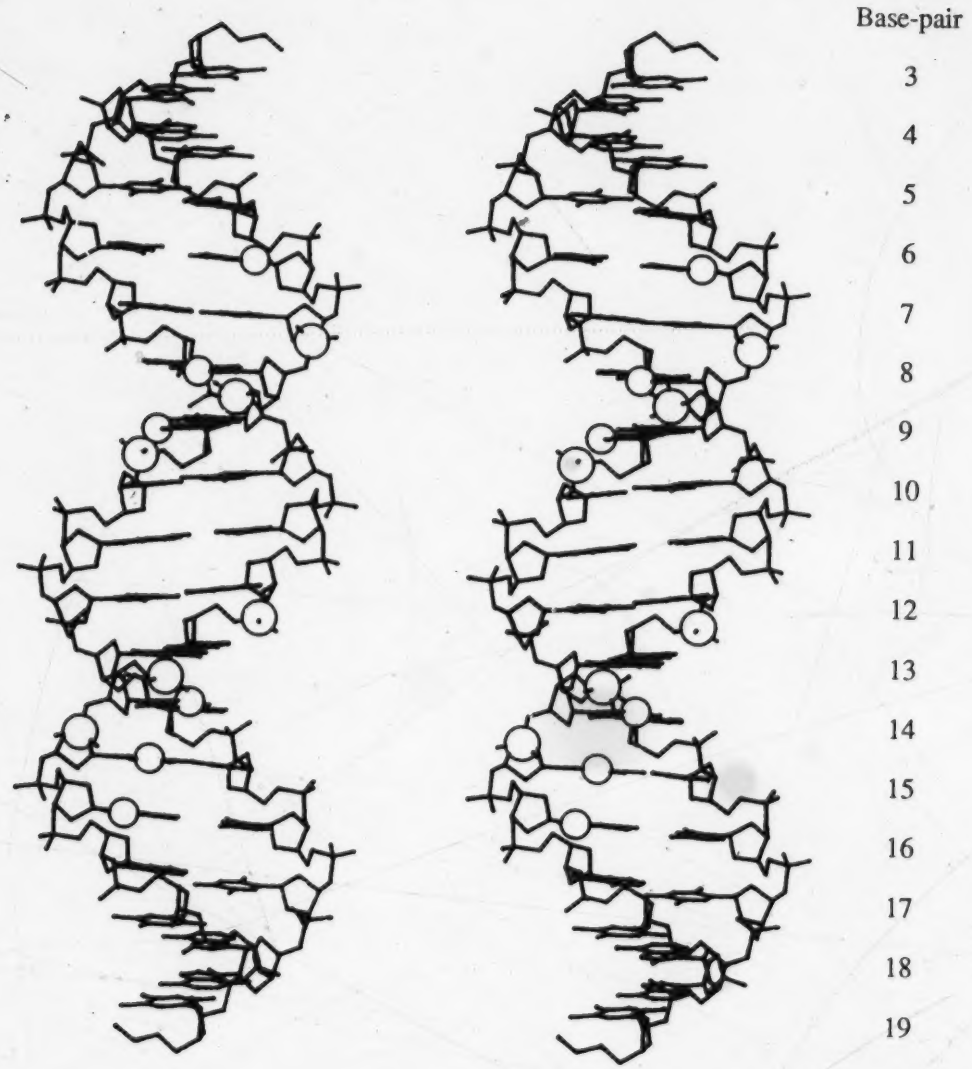


Figure I-7 The O_R3 operator in a B DNA conformation

A stereoview of the 17 base-pair O_R3 operator sequence in a classical B DNA conformation. Phosphates that, when ethylated, interfere with Cro repressor binding are shown with large spheres. Smaller spheres indicate the nitrogens protected by Cro from methylation with dimethylsulfate. Numbers indicate base-pair position relative to the 21 base-pair operator.

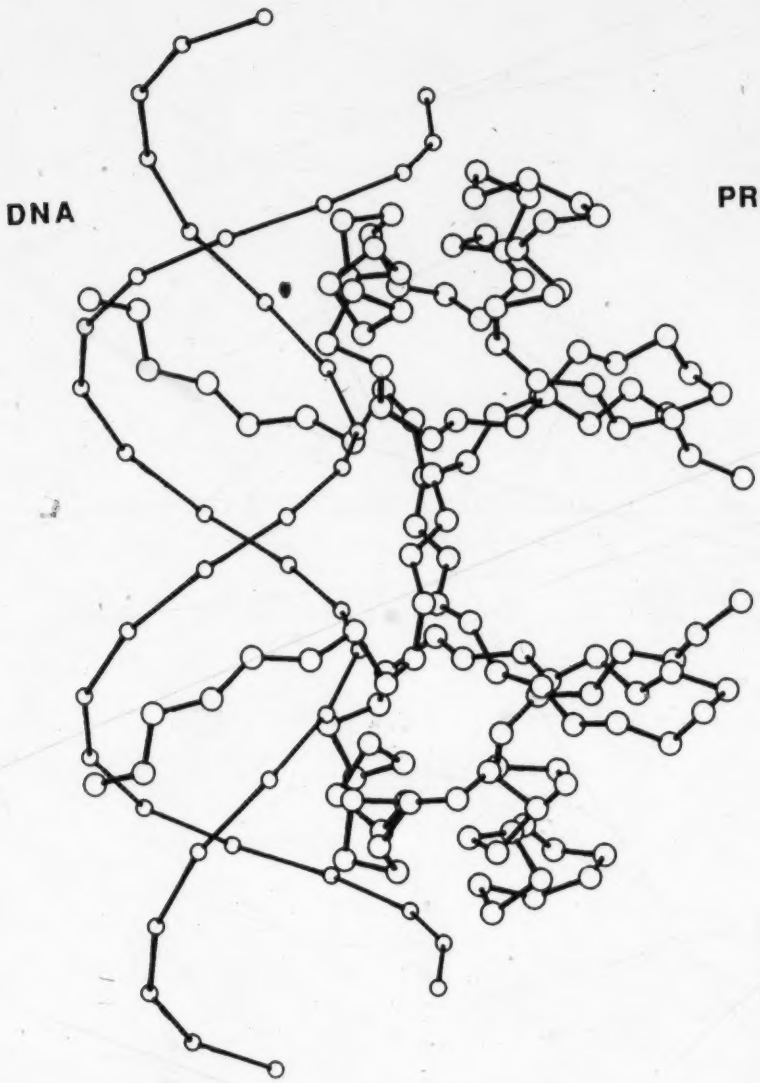


Figure I-8 Model of Cro-DNA interaction

A model of the interaction of Cro repressor protein with operator DNA was built using the crystallographically determined structure of the unbound protein. The protein was placed so that its two-fold symmetry axis was coincident with the pseudo-two-fold axis of the operator DNA which was in a classical B type DNA conformation. The distance between protein and DNA, their relative angle about the common two-fold axis, the rise and twist per base-pair, and curvature of the DNA were allowed to vary with the conformation of the protein side-chains to optimize interactions in the complex. Only DNA phosphates and protein α-carbon atoms are shown.

not driven merely by long-range electrostatic interaction between the positively charged protein and negatively charged DNA since the extent of the acceleration decreases markedly with smaller DNA fragments (less than 75 base-pairs). Cro may exist in a distinct open conformation when interacting with non-specific DNA. Since the non-specific DNA-Cro complex is salt-sensitive (Boschelli, 1982), most of these interactions are ionic in nature[2]. The open conformation would allow some freedom in sampling base-pair atoms in the major groove (Ohlendorf et al., 1982) as it rapidly 'slides' along the DNA in an assisted or one-dimensional diffusional process in search of the operator sequence.

Since both Cro and the DNA binding sites have approximate two fold symmetry, it is initially sufficient to consider the interaction between one monomer of Cro and half of the binding site. The proposed scheme for sequence-specific binding of Cro to DNA is presented in Figure I-9. The architecture of the protein reveals that the recognition helix is held by the rest of the protein in such a position as to read the hydrogen bond donating, hydrogen bond accepting, and van der Waals contacts presented by B DNA. Any deviation from this pattern possibly results in the loss of interaction between protein and DNA. Although this model could explain much of the chemical modification data and the magnitude of the affinity of Cro for naturally occurring mutants of the O_R3 operator DNA, a number of assumptions had been made in its construction. The model allowed for only small local deviations from regular B DNA, whereas other crystal structures of DNA, even in the absence of protein, had shown more heterogeneity within their structure, for example, in the base-stacking between adjacent base-pairs (Dickerson & Drew, 1981). Also, a change in the relative positioning of the two monomers in the Cro dimer (a hinge-opening motion) was noted to be essentially equivalent as DNA bending in achieving optimal interaction between protein and DNA.

ii. Testing the proposed Cro-DNA model

A large number of papers have tested the model of the Cro-DNA complex proposed by Ohlendorf et al. (1982) on the basis of the protein structure alone. Here, I evaluate and review the most systematic (Sarai & Takeda, 1987; Takeda et al., 1989; Caruthers et al., 1986; Pakula et al., 1986, Takeda et al., 1986; Eisenbeis et al., 1985), the most technically interesting (Kim et al., 1987; Kirpichnikov et al., 1985; Shirakawa et al., 1985; Metzler &

---

[2]The stability of the specific cro-DNA complex is relatively insensitive to ionic strength of the media, consistent with most of the interactions between protein and DNA being hydrophobic and hydrogen-bond contacts (Boschelli et al., 1982).

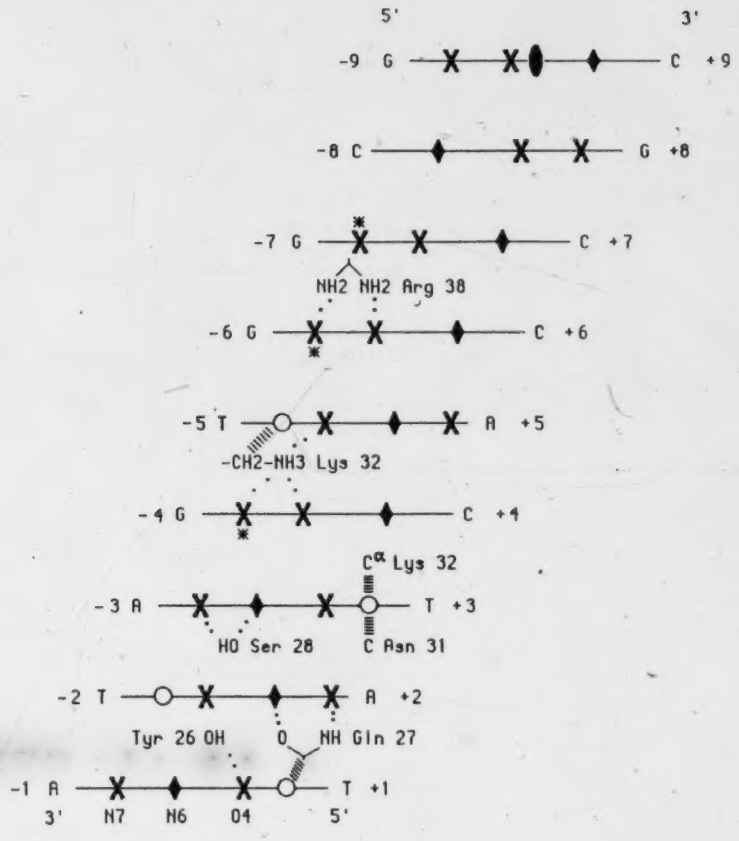


Figure I-9 Sequence-specific interaction between DNA and Cro repressor

This model (adapted from Ohlendorf et al., 1982) shows a schematic representation of the sequence-specific interactions between the amino acid side-chains of Cro and the edges of base-pairs exposed in the major groove of DNA. The view is approximately into the major groove with the base-pairs seen edge-on. The center of two-fold symmetry for the full 17 base-pair operator DNA sequence is indicated by the dyad symbol at base-pair 9. The sequence from +1 to +9 corresponds to base sequence 3 to 11 in Figure I-5 for the O_R3 and consensus operators (and the sequence from -9 to -1 corresponds to base sequence 32 to 40). The symbols are: X , a hydrogen-bond acceptor; ♦ , a hydrogen-bond donor; O , the methyl group of thymine; • • , presumed hydrogen bond; |||||||||| , van der Waals contact to a thymine methyl group. Guanine N7 nitrogens that are protected from methylation when Cro is bound are indicated by an asterisk.

Lu, 1989), and a preliminary report of the crystallographic determination of the Cro-consensus operator complex (Brennan & Matthews, 1989a). All support the general features of this model, and all suggest refinements in specific details.

One of the first reports supporting the model came from the discovery of quite striking amino acid and structural homology in the region including the second and third helices of Cro with other DNA-binding proteins (reviewed by Brennan & Matthews, 1989b). The common sequence is now referred to as the helix-turn-helix motif. The first helix of the motif (α2 of Cro) position the second helix (α3 of Cro) to be in a correct orientation for reading the contacts presented in the major groove of DNA. The sequence of Cro is compared to that of related proteins in Figure I-10. Only the λ repressor, the *E. coli* Lac repressor, the *E. coli* Trp repressor, the phage 22 repressor, and the repressor and Cro proteins from bacteriophage 434 are included (and are discussed further below). Homology has been noted for many other proteins, including those of eukaryotic organisms (Brennan & Matthews, 1989b). Sequence homologies include the invariant glycine required in the turn between the helices, a glutamine at position 1, and hydrophobic amino acids at defined positions in the motif to stabilize the relative positioning of the helices in the helix-turn-helix.

			<——— Helix ———>	<-Turn->	<——— Helix ———>																		
		-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
λ Cro	14	F	G	Q	T	K	T	A	K	D	L	G	V	Y	Q	S	A	I	N	K	A	I	H
λ Rep.	31	L	S	Q	E	S	V	A	D	K	M	G	M	G	Q	S	G	V	G	A	L	F	N
434 Cro	16	M	T	Q	T	E	L	A	T	K	A	G	V	K	Q	Q	S	I	Q	L	I	E	A
434 Rep.	16	N	L	Q	A	E	L	A	Q	K	V	G	T	T	Q	Q	S	I	E	Q	L	E	N
Trp Rep.	66	M	S	Q	R	E	L	K	N	E	L	G	A	G	I	A	T	I	T	R	G	S	N
Lac Rep.	167	V	T	L	Y	D	V	A	E	Y	A	G	V	S	Y	Q	T	V	S	R	V	V	N
P22 Rep.	23	R	G	Q	R	K	V	A	D	A	L	G	I	N	E	S	Q	I	S	R	W	K	G

Figure I-10 Sequence homologies with Cro repressor

The helix-turn-helix amino acid sequence for a few representative DNA-binding proteins are aligned by sequence homology. Amino acids that are most strongly conserved (A or G in position 5, G in position 9, and I or V in position 15) are underlined. The numbers at the top follow the convention of Pabo & Sauer (1984). The number to the left of each partial sequence is the amino acid sequence number of the first residue in the segment (Adapted from Brennan & Matthews, 1989a).

A circular dichroism (CD) study examined the conformation of operator when free and when bound to Cro repressor (Kirpichnikov et al., 1985). In the absence of protein, the O_R3 operator displays a positive CD band at 270 nm. and a negative one at 250 nm., which is typical for classical B DNA. Addition of Cro repressor suppresses the positive band, with the most pronounced effect at 280 nanometers. Such a spectral change is quite distinct from that of a B to A DNA transition, or in non-cooperative winding or unwinding of DNA[3]. The CD change was inferred to be the result of an induced CD in tyrosine 26 (Figure I-11) upon its interaction with a specific base-pair in the operator. The operator conformation is conserved in a B-like form as a whole, although local distortions, such as kinks could not be discounted from the CD data alone. In summary, the operator DNA exists overall in a B DNA conformation either as a free duplex, or in complex with the Cro protein, validating the use of a B DNA conformation in the proposed model.

Site-directed mutagenesis has been applied to many of the amino acid residues of Cro, and generated mutants have been examined *in vitro* (Eisenbeis et al., 1985; Caruthers et al., 1986) and *in vivo* (Pakula et al., 1986). Generally, altering any of the amino acid residues proposed to make contact in the major groove of DNA (Figure I-9) reduces the specificity for operator DNA, and often, overall DNA affinity as well. For example, replacements of Gln 27 for Leu, Cys, Arg, or His, results in reduced Cro binding to O_R3, but at least in the case of the Gln to His substitution does not appreciably alter protein structure. Replacement of Ala 33 with Lys, Arg, Thr, Val, and Gly, also reduces the ability of Cro to bind to O_R3, despite Ala 33 not being predicted to make direct contact to the DNA. Here, however, Ala 33 resides on the hydrophobic side of the recognition helix, and substituting any group may be expected to have a disruptive effect on Cro structure, and therefore DNA binding. Other amino acid changes, such as Ala 36 to Ser, Thr, Lys, Arg, or Gln do not appreciably change the affinity of Cro for DNA, consistent with the role of Ala 36 not packing into the hydrophobic core of the protein, nor predicted to make contact with DNA. The Val 55 to Cys mutation will be discussed in more detail later (Chapter V). The results support the current model for the interaction of Cro repressor with operator DNA.

The role of tyrosines and lysines in Cro has been examined by chemical modification studies. In the absence of DNA, all eight lysines and the terminal amino group of Cro can be alkylated, and all three tyrosine residues can be iodinated. In the presence of 21 base-pair operator DNA, lysines 32 and 56, and all three tyrosines (10, 26 and 51) are fully protected from modification (Takeda et al., 1986), and lysines 21, 62, and 63 are partially

---

[3]As is observed in the complex of DNA with the Lac repressor and with the catabolite gene activator protein (CAP).

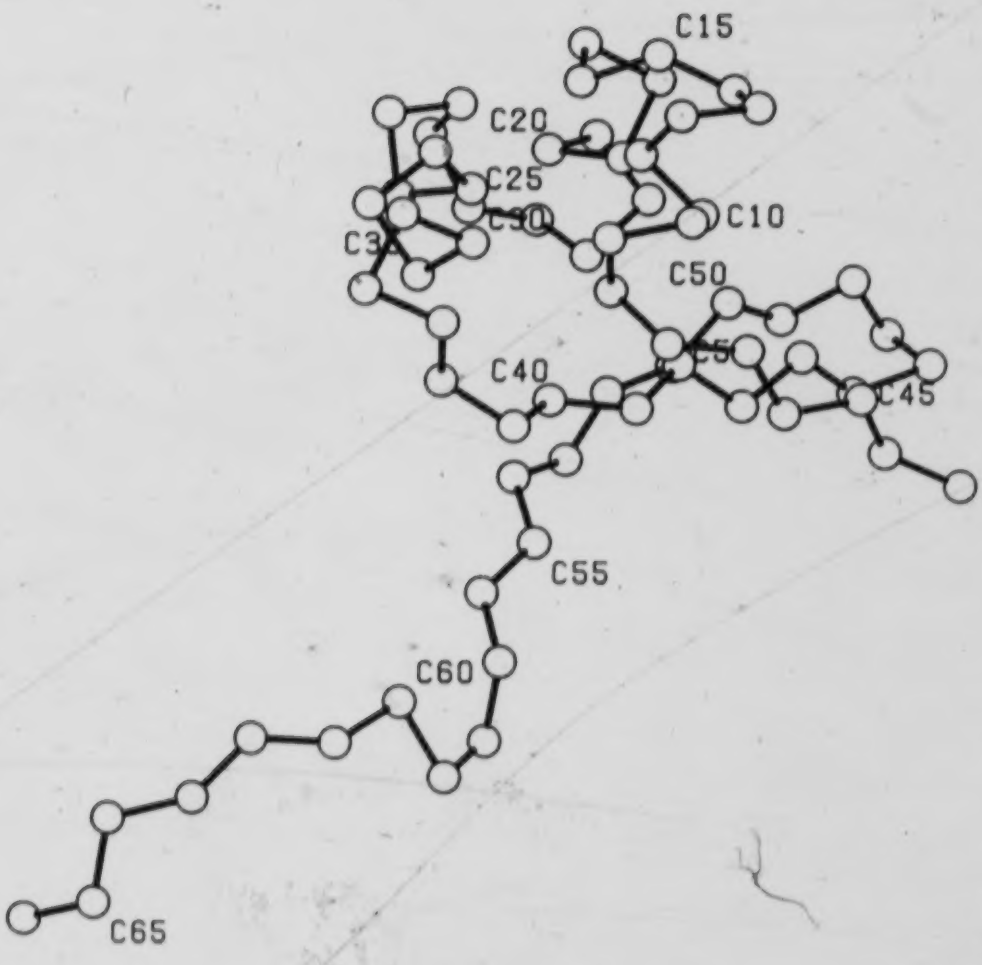


Figure I-11 Cro repressor protein

An α carbon representation of one monomer of Cro repressor in the same orientation as in Figure I-8. Every fifth residue is labelled with the amino acid sequence number.

protected. The protected amino acid residues are generally located on one side of the Cro dimer, defining a DNA-binding region that is consistent with the proposed model.

The chemical modification data on Cro repressor also gave experimental support for the proposed model with non-specific DNA. An important feature of the non-specific complex is flexibility about the hinge portion of Cro (residues 54-56) and of the C-terminal arms. Such mobility may allow each of the Cro monomers to be in loose association with DNA and allows for rapid scanning of the DNA sequence. In going from the non-specific to the specific complex, the Cro repressor likely moves closer to the DNA and penetrates more deeply into the major groove. As this occurs, close contacts develop between the DNA backbone at two or three base-pairs out from the center of the operator and residues 57-60 at the base of the flexible carboxy terminal arms. To relieve this tight complex, the arms twist out of the way, which accounts for increased exposure of lysines 62 and 63 to modification in the specific complex over the non-specific one. Iodination data for the Cro specific DNA complex also supports a different Cro conformation in the sequence-independent interaction with DNA. Tyrosine 26 is susceptible to modification in the non-specific complex, which is consistent with the α3 helix further removed from the DNA.

The conformation of Cro in the non-specific DNA-Cro complex is however different to that of the apo protein. Tyrosine 10 remains completely protected, and tyrosine 51 partially protected from iodination. The protection with tyrosine 10 is not because of direct contact with DNA, but might be due to a conformational change in the protein upon DNA binding.

Chemically induced dynamic nuclear polarization (CIDNP) ^{1}H NMR experiments measure the accessibility of aromatic side chains to a flavin dye present in solution. It is therefore interesting to compare results from the iodination chemical probe experiments to that of the CIDNP study of Cro (Shirakawa et al., 1985). In the absence of DNA, the signal due to the 3,5 protons of tyrosine 26 gives rise to a remarkable emission doublet and tyrosine 51 3,5 protons have a weaker CIDNP intensity, whereas tyrosine 10 does not have any CIDNP effect. This is consistent with the crystal structure of the free Cro, and with mild nitration modification experiments (Shirakawa et al., 1985), in which tyrosine 26 is fully exposed to solvent, tyrosine 51 is substantially exposed, and tyrosine 10 is buried within the protein. The iodination of tyrosine 10 in the free protein noted above suggests that the iodination reaction is quite vigorous, as this residue cannot be modified by mild nitration, and is not accessible to a flavin dye in the solvent. The addition of O_R3 17mer operator DNA or non-specific DNA reduces the CIDNP intensity for all peaks, i.e. including tyrosine 26, suggesting that it is not as completely exposed as was suggested by the iodination modification experiment.

^{19}F NMR has also been applied to the study of O_R3-Cro interaction (Metzler & Lu, 1989). Each thymine of the O_R3 operator was replaced selectively with 5-fluorodeoxyuridine, which alters little the affinity of Cro for binding DNA (Metzler et al., 1985). Chemical shift changes upon addition of Cro result from either direct contact with the protein, or from an induced conformational change in the 17mer O_R3 operator DNA (which corresponds to base-pairs 3-19 of the 21 base-pair operator, Figure I-5). The fluorine of base 36, undergoes a large (1.2 ppm) upfield change in chemical shift, consistent with the methylene groups of Lys 32 of the protein making a van der Waals contact to the fluorine of this base (and by analogy, to the methyl group for thymine 36 of the unsubstituted operator).

The methylene protons of Gln 27 are predicted to interact with the methyl group at thymines 3 and 24. Although fluorine 3 does not undergo any chemical shift change, its symmetry related partner fluorine 24 does. Fluorine 3 (methyl group of thymine 3 in the unsubstituted DNA) is in the left half of the O_R3 operator, which contains the consensus half-operator sequence, and should therefore have the proposed Cro contact. The absence of a chemical shift change at fluorine 3 suggests that either the model is wrong at this point, or that a change in the conformation of DNA produces an equal and opposite change in the chemical shift. The differences between the two ends is likely not due to a difference in the conformation of DNA in the absence of protein since: (1), the last four pairs of the operator are symmetric; (2), the fluorines 3 and 24 have nearly identical chemical shifts (-89.1 ± 0.1 ppm, other fluorines are between -87.8 and -88 ppm); and (3), in the unsubstituted 17mer, the ^1H resonance positions of thymine residues 3 and 24 are the same (average ± 0.03 ppm). In the complex, there is an asymmetry in environment at the two ends of the operator (and is discussed further in Chapter VI).

Fluorines at positions 5 and 26 are also predicted to make van der Waals contact with the side chains of residues Asn 31 and Lys 32 of Cro. However, these undergo a substantial downfield shift, suggesting a change in operator conformation such as local unwinding of the DNA helix. Chemical shift changes are also observed with fluorines on base 30 and 31, near the center of the operator, which are not contacted by Cro. This suggest that the sequence-dependent contacts with the region of base pairs 3-9 (and the symmetry related region) induces a change in DNA environment in the middle of the operator.

Solvent induced isotope shifts (SIIS values[4]) were used to probe the accessibility of each of the fluorines in the free operator, and in the protein-operator complex. In the

---

[4] SIIS = δ (90% H_2O/10%D_2O) - δ (100% D_2O)

absence of protein, all fluorines have about the same SIIS values, approximately equal to 0.15, implying a rather regular DNA structure. The value for free 5-fluorodeoxyuridine is 0.39 ppm, which indicates that the fluorines in DNA are shielded from solvent, even in the absence of protein.

In the operator-protein complex, fluorine 36 becomes nearly completely inaccessible to solvent (SIIS ≈ 0.0), consistent with an interaction with lysine 32 of Cro. Fluorines 5 and 26 have larger SIIS values (0.26), indicating increased exposure to solvent, and local unravelling at these positions upon binding Cro. The solvent accessibilities of the other fluorines undergo little change on the addition of Cro repressor.

One of the most convincing experiments supporting the proposed model (Ohlendorf et al., 1982) for the Cro-DNA complex comes from the systematic base substitution experiments of Takeda et al. (1989) and the preliminary crystallographic structure of the Cro-O_R3 17mer complex.

In the base substitution experiment, the binding affinities of purified Cro repressor were quantitatively measured for chemically synthesized wild type and sixty-eight (68) different mutant O_R1 operators. Each mutant operator was different from wild type O_R1 in a single base-pair. All base pairs were systematically changed to either of all three possible base pair substitutions, or the thymine was changed to a uracil (Figure I-12) . Mapping the affinity changes along the 17 base-pair operator clearly shows where the sequence-specific interaction occurs within the operator and how each base or base-pair energetically contributes to the specific binding (Takeda et al., 1989). The binding free energy changes calculated from the observed change in affinity are mostly additive for specific Cro binding. The binding affinities of Cro to the operator DNA sequences indicated in Figure I-5, or any other DNA sequences, can be predicted accurately by simple addition of free energy changes of single base-substitutions. Figure I-13 shows a correlation plot between the predicted and measured free energy changes for the six operators and consensus operator DNA sequences in the regulatory region of phage λ for which numerical data were available (Takeda et al., 1989; Sarai & Takeda, 1987; Kim et al., 1987). The complete plot (Figure 2 of Takeda et al., 1989) shows a striking correlation for changes in free energy of up to + 5 Kcal/mole, or a 10^4 loss in binding affinity (and includes non-specific DNA binding).

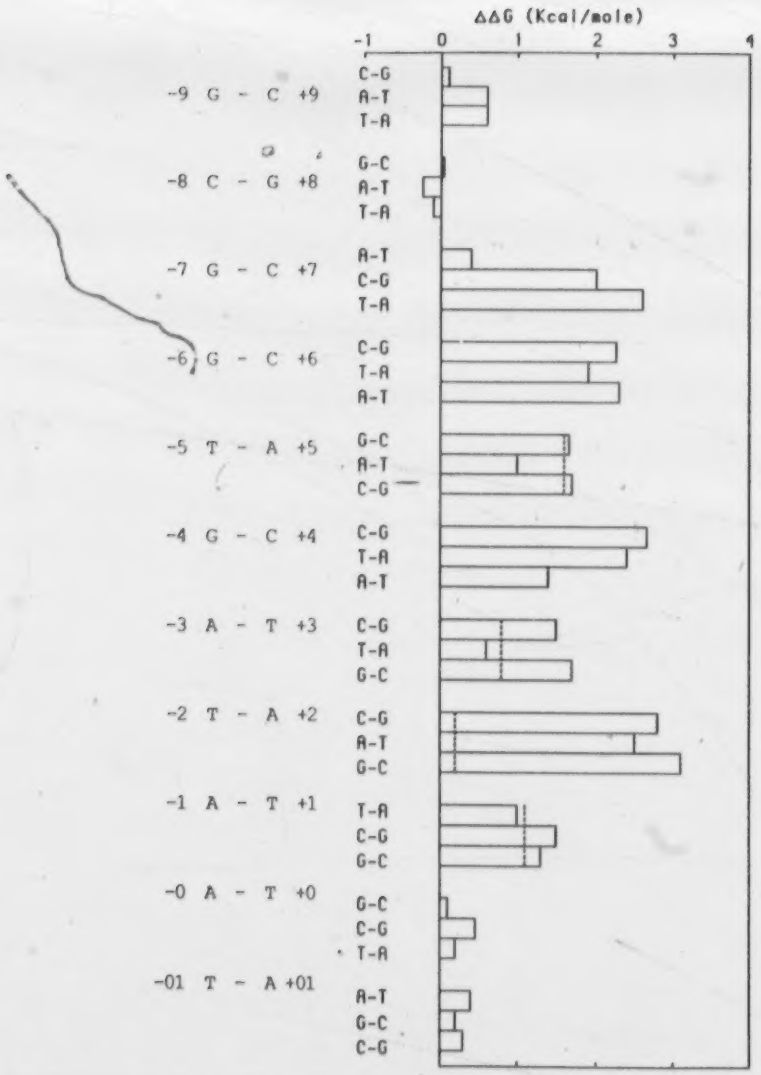


Figure I-12 Effect of base substitutions on the binding of Cro to operator DNA

Binding free energy changes are shown for each of the base and base-pair substitutions in the consensus half (right half in Figure I-5) of the O_R1 operator (Adapted from Sarai & Takeda, 1987). The free energy change is calculated from the ratio of equilibrium dissociation constants ($\Delta\Delta G = -0.546 \ln (K_d \text{ (subst.)} / K_d (O_R1))$ as measured by a filter-binding assay (Sarai & Takeda, 1987; Takeda et al., 1989). Bars indicate the free energy changes due to the base-pair substitutions. A dashed line indicates the change when a thymine of O_R1 is replaced by uracil. The changes for the non-consensus half of the operator are very similar (Sarai & Takeda, 1989).

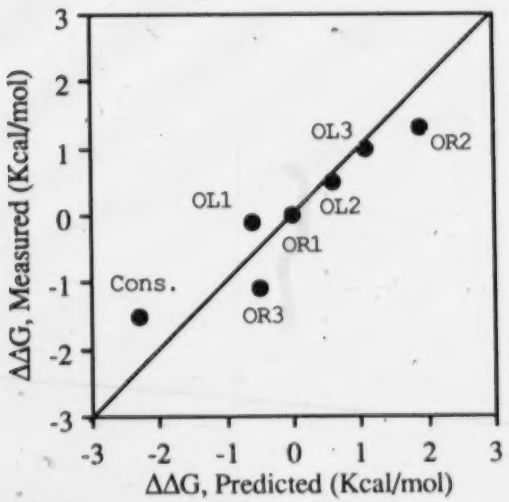


Figure I-13 Prediction of the binding affinity of Cro for DNA

This figure shows a correlation plot between measured binding free energy changes and those predicted from the base-substitution experiments for the six regulatory operator DNA sequences and the consensus operator sequence (Cons.) derived from them. The 45° line indicates perfect agreement between the change in free energy predicted and that measured. Changes are relative to the binding affinity for the O_R1 operator. All operators were synthesized as duplexes 21 base-pairs long (adapted from Takeda et al., 1989).

The sequence-specificity of Cro lies in base pairs 1 to 7 of Figure I-9. Substitutions in the center of the operator, or outside this sequence do not appreciably alter the affinity of the DNA for Cro repressor. The systematic base substitution experiments have also indicated subtle changes to the model proposed by Ohlendorf et al. (1982). For example, the guanine at position -7 (Figure I-9) is protected from methylation when Cro is bound, although no hydrogen bond was predicted to be made to the N7 at this position. In the original model the protein was assumed to sterically inhibit access by dimethyl sulfate to this base. However, replacement of the G:C base-pair by any other results in a substantial loss in DNA affinity. With additional model building, Takeda et al. (1989) invoke a revised model in which Arg 38 also contacts the guanine at -7 (Figure I-9). The thymine to uracil replacement confirmed the importance of van der Waals contacts to the Cro protein made by the DNA. All thymine methyls are involved, except Thy -2. Confirmation of either the original or revised models for the interaction of Cro and DNA requires high-resolution structural studies either by crystallographic means, or by NMR spectroscopy.

Brennan & Matthews (1989b) have published preliminary results from the X-ray diffraction study of a Cro-17mer consensus operator at moderate resolution. The recognition helix of Cro (α3) lies within the major groove, and the side chains of this helix contact the exposed edges of the DNA bases. The operator is bent, with the DNA in the middle overwound, and the minor groove somewhat compressed. Although the 3-dimensional structure of the complexes and uncomplexed monomers are the same, the Cro dimer undergoes a large conformation change upon binding to DNA. The relative positioning of the monomers changes dramatically by twisting about the β3 strands such that the helix-turn-helix motifs of each monomer are rotated by more than 35° with respect to each other. Higher resolution will be required to visualize details of protein side-chain to DNA base-pair interactions.

In summary, the role of the Cro protein *in vivo* is to locate and bind to the O_R3 operator DNA sequence in bacteriophage λ. This binding represses synthesis of the product of the *cI* gene, λ repressor, and commits the phage to lytic mode of growth. The two recognition helices of Cro are positioned to interact within two successive major grooves of DNA. The protein distinguishes the O_R3 sequence from the others by using the recognition helix to read the pattern of hydrogen-bond acceptors, hydrogen bond donors, and van der Waals contacts presented in the major groove of the DNA.

The remainder of this introduction is devoted to comparing the features of other protein-DNA complexes to the Cro-DNA complex.

d. The λ repressor from bacteriophage λ

The λ repressor has been found to have a DNA-binding substructure related to that of Cro, which is not surprising since they both regulate gene expression by binding to the same six sites in phage DNA (although with differing affinities). The λ repressor is also dimeric, but here, each monomer consists of a two-domain structure. The C-terminal domain contains most of the monomer-monomer contacts stabilizing the DNA-binding dimer structure. Cleavage within this domain (e.g., that assisted by RecA protein during prophage induction) considerably weakens the monomer-monomer interaction and therefore the concentration of the active dimeric form of the protein needed for efficient DNA-binding under normal cellular conditions.

The N-terminal domain of λ repressor, residues 1 to 92, is the DNA-binding domain which makes all of the contacts to DNA required for control of transcription (Ptashne et al., 1980). The crystal structure of this domain (Pabo & Lewis, 1982) shows an all-helical protein with a helix 5 extending out from the globular monomer domain to contact helix 5'

of the other subunit to form a dimer interface. The helix-turn-helix motif was found to be quite similar to that of Cro repressor, with an rms deviation for 21 α carbons of 0.7 Å (Brennan & Matthews, 1989b). Although the recognition helices are 34 Å apart, they are not tilted to the degree seen in Cro. Building a model of the λ repressor-DNA complex in which the recognition helices are fit into two successive major grooves of B DNA places the N-terminal part of this helix more deeply into the major groove than the rest of the recognition helix. Therefore, although the two proteins, Cro and λ repressor, share a structural motif, they do not necessarily use it in precisely the same manner. In addition, the λ repressor has a flexible N-terminal arms that wraps around the DNA and makes sequence-specific contact in the major groove on the 'back-side' of the DNA. By way of contrast, the C-terminal arm of Cro makes non-specific contacts in the minor groove.

Sarai & Takeda (1987) have also measured the affinity of λ repressor to the O_R1 operator and a series of systematic base substitution mutants of this operator. Analogous to the experiments with Cro, each of the base-pairs in the consensus half of the operator (base-pairs 11-19, Figure I-5) was replaced, in turn, by each of the possible three base-pair substitutions, and each thymine was replaced by uracil. The base-substitution experiments predicted that the thymine 15 was most important to make a van der Waals contact with the protein, thymine 24 was somewhat less important, and the methyl groups of thymines 18 and 26 not required for λ repressor binding. The adenine at base-pair 18 was important, as in Cro, but the identity of base-pairs 16 and 17 less important. Sequence-dependent interactions are made with base-pairs 12 to 14. Strong interactions are concentrated around the center of the operator for λ repressor, where the N-terminal arm is modelled to make a sequence-specific contact. In Cro, the identity of the central base-pairs of the operator DNA does not greatly influence protein binding.

The crystallographic determination of the λ repressor-O_L1 operator site at 2.5 Å resolution supports the general features of the model which fits the protein dimer against linear B form DNA (Figure I-14). The complex is stabilized by an extensive network of hydrogen bonds between the protein and the sugar-phosphate backbone. Specificity of binding is provided by amino acid side chain hydrogen bonding and van der Waals contacts with sites in the major groove of DNA. However, the cocrystal structure reveals important side chain-side chain interactions not predicted from the modeling or from previous genetic and biochemical studies. Since most of the contacts are approximately the same in the two halves of the operator site, interactions with only the consensus or right half of the DNA are discussed. These are summarized in Figure I-15. On the left side of the figure, the interactions are shown from the original model of Pabo & Sauer (1984), and on the right, are the contacts inferred from the crystal structure of the complex. It should be pointed out

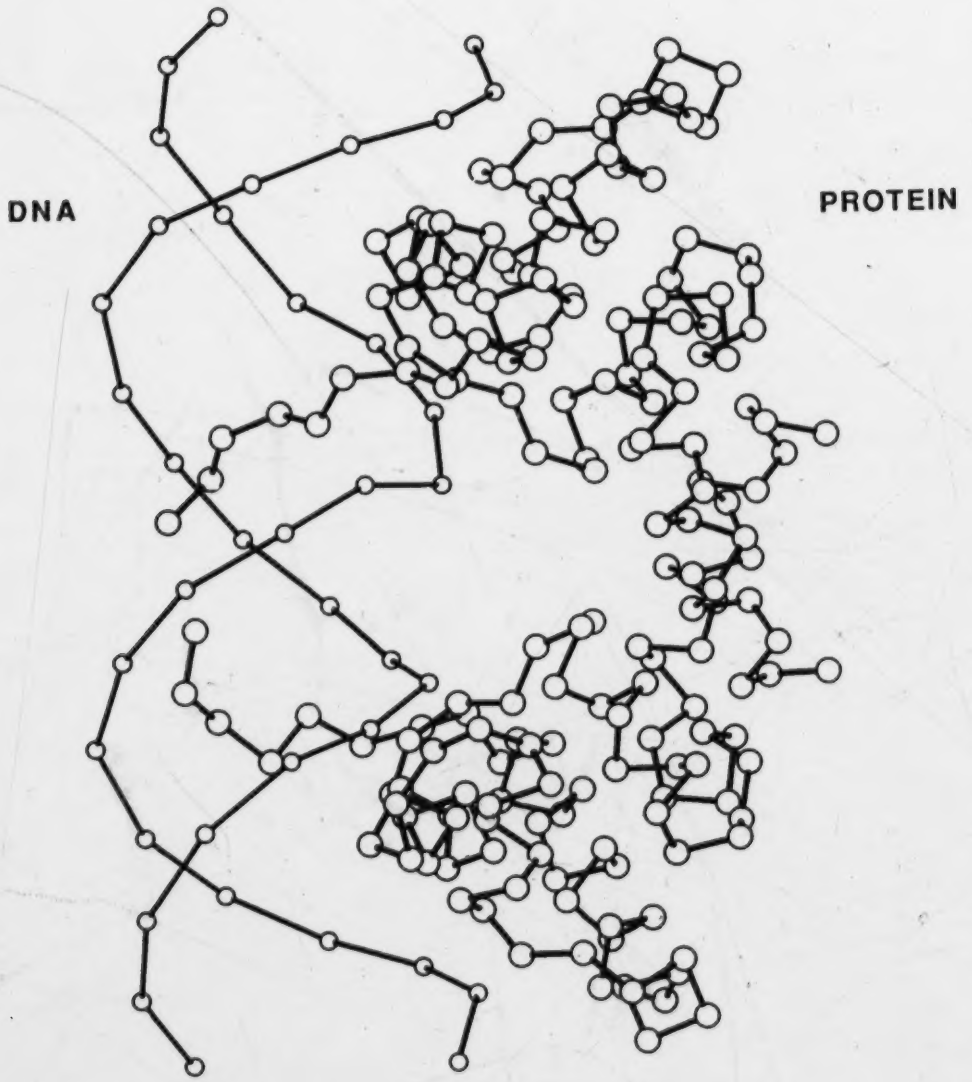


Figure I-14 Structure of the λ repressor-DNA complex

The structure of the λ repressor N-terminal domain-operator DNA complex as determined crystallographically (Jordan & Pabo, 1988). Only phosphate atoms of the DNA and α carbons of the protein are shown for clarity. The N-terminal arm of one of the monomers showed poor electron density and was omitted from structure determination. The view is approximately the same as the model presented for Cro-DNA interaction (Figure I-8).

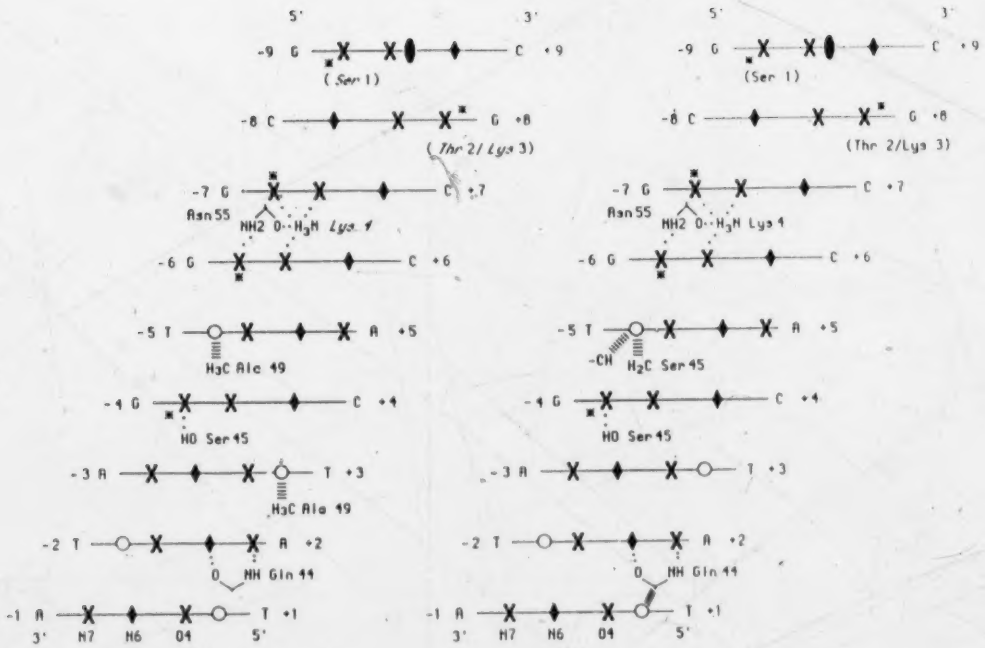


Figure I-15 Sequence-specific interactions between DNA and λ repressor

The left part of the figure shows the prediction of sequence-specific interaction between λ repressor protein and operator DNA based on the model of Cro repressor interaction with DNA (from Pabo & Sauer, 1984). Contacts made by amino acid residues 1-4 are copied from the right side, as no prediction was originally made for their interactions. The right side shows the actual contacts observed in the crystal structure of the λ repressor N-terminal domain-operator DNA complex, as interpreted by Sarai & Takeda (1989). Only interactions in the consensus part of the operator are shown. Tentative interactions are in parentheses. Diagram symbols are explained in the legend to Figure I-9.

that even at 2.5 Å resolution, the X-ray structure can be subject to interpretation. For example, additional contacts by Lys 4 to Gua -7 (in addition to -6) are invoked by Sarai & Takeda (1989) to fully explain their systematic base-substitution experiments. In this work, they extend the base-substitution experiments to all base-pairs of the O_R1 operator. Although the λ repressor dimer is comprised of identical subunits, the protein recognizes the pseudo-two-fold symmetric operator site asymmetrically. Base substitutions within the consensus half of the operator affect binding more than base substitutions within the non-consensus half of the operator. Changing a non-consensus base-pair to the consensus base-pair does not necessarily increase, and often decreases binding.

The crystal structure shows that the relative positioning of the amino acids to the DNA bases is slightly different in the two halves of the protein. This asymmetric binding is proposed to be the result of steric clash between two N-termini at the center of the dimer, and subsequent dislocation of one of the monomers when both attempt to simultaneously bind to the operator symmetrically. As in the Cro-DNA system, λ repressor directly reads the DNA sequence by forming specific hydrogen bonds and hydrophobic contacts to the DNA bases. Monomers of both repressor proteins bind strongest to the consensus half-operator. However, unlike Cro, if both halves of the operator are the consensus sequence, λ repressor affinity is reduced relative to the O_R1 sequence. It appears that the purpose of a consensus sequence is to bind one-half of a dimeric protein strongly to one-half of the DNA site in the phage λ regulatory region—discrimination between DNA sequences actually occurs with the non-consensus, more weakly interacting site.

2. DNA recognition by the Cro and repressor proteins of bacteriophage 434

The repressor and Cro proteins of temperate phage 434, a close relative of phage λ, bind to a set of six related fourteen base-pair operator sites in the phage genome. The binding affinity for the two proteins differ from site to site, and, as in λ, this differential binding creates the regulatory switch between lytic and lysogenic modes of life. 434 repressor is similar to λ repressor in that it has two domains, with the N-terminal 69 residues forming a DNA-binding domain. Both the N-terminal domain of 434 repressor ($R_{1\text{-}69}$) and the 434 Cro repressor have have their crystal structures determined (Mondragón et al., 1989a; Mondragón et al., 1989b). They each contain the DNA-binding helix-turn-helix motifs analogous to that of Cro. The rms deviation from λ Cro (for the 21 α carbons of the motif) is 0.43 Å for 434 repressor, and 0.56 Å for 434 Cro. The sequence and structural homology between the 434 repressor and 434 Cro is much greater than that

between the similar proteins of λ phage. 48% of the amino acid residues were identical, and backbone atoms of the subunits superimpose to 0.77 Å.

In order to compare how the two proteins from phage 434 bound DNA, co-crystals were prepared with the same 14 base-pair sequence (Wolberger et al., 1988). It was anticipated that they would form very similar complexes when bound to the same DNA fragment. The surprising finding is that the same 14 base-pair operator adopts a different conformation, depending on whether 434 Cro or 434 R_{1-69} is bound. In both complexes, the proteins are bound to DNA as a dimer with the recognition helix of each monomer lying in successive major grooves of the DNA. The differences between the complexes (reviewed in Wolberger et al., 1988) can be divided into three categories: (1), the relative disposition of the two subunits in each protein dimer; (2), the conformation of the operator; and (3), the relationship between protein and DNA in each operator half-site.

The side chain at the dimer interface differs in size between the two proteins, and therefore how the monomers pack. As a result, the relative orientations of monomers in the two complexes are not the same. In the complex with operator, the relative orientation of two 434 Cro monomers is consistent with essentially straight DNA. The different orientation of the monomers in 434 R_{1-69} imposes a bend. As a result, different contacts can be made by 434 Cro and 434 repressor.

It is conceivable that there could be two proteins containing identical DNA recognition helices in each monomer, but with differing dimer interface regions. Because the two monomers could placed differently, these proteins could recognize unequal DNA sequences. Although it may be possible to devise a set of recognition rules for a given DNA-binding protein, they will not generally be true, even for proteins that recognize DNA with the same structural motif. It remains a difficult task to predict the cognate DNA sequence from the amino acid sequence or even the structure of a DNA-binding protein. A given operator sequence is not recognized by a simple set of rules in which certain amino acids, even in the helix-turn-helix motif, confer specificity for certain bases.

The crystallographic study of a phage 434 repressor-20 base pair DNA complex at high resolution (Aggarwal et al., 1988) has revealed unique aspects of operator deformation upon forming the protein-DNA complex. From the lower resolution determination of the structure of the 434 repressor-operator complex, the recognition α-helix of the repressor lies in each half-site of the operator (Anderson et al., 1987). Each of the helices is positioned so that contacts are made to the outermost five base-pairs on each half-site of the 14 base-pair operator, but not to the innermost base-pairs of the operator. However, the sequence of the central base-pairs of the operator have a significant role in determining operator affinity for repressor—operators bearing A:T or T:A base-pairs in position 6 to 9

bind 434 repressor more strongly than does operator with G:C or C:G base-pairs at these positions (Koudelka et al., 1987). The base composition near the center of the operator was proposed to affect affinity of the DNA for 434 repressor not by altering DNA conformation in the absence of protein, but by altering the ease with which operator DNA can be deformed into the optimal configuration for complex formation. Evidently this DNA deformation is required to align the two operator half-sites so that each monomer of the bound repressor dimer can make optimal contacts with each half-site.

The preliminary low resolution (7 Å) X-ray study had shown that the DNA near the center is over-twisted, so that the minor groove is compressed (Anderson et al., 1987). The higher resolution 2.5 Å study reveals that the overwinding near the center causes the central base-pair to assume a non-coplanar conformation. The central A:T base-pairs are sufficiently propellor twisted to form 3-centered bifurcated hydrogen-bonds, and stabilize the conformation of the bound DNA. The extra torsional rigidity imparted by the G+C-rich sequences as a result of the additional base-pair hydrogen-bond can account for the observed decrease in affinity (Koudelka et al., 1988). When fixing a particular conformation for the DNA, the protein appears to meet less resistance from one sequence than from some other. In the interaction between 434 repressor and its operator, the protein appears to position the DNA backbone, with a cost in free energy that depends on base sequence. The presence of hydrogen bonds additional to normal Watson-Crick base-pairing is an important component of this adjustment.

A number of important features of protein-DNA recognition has emerged from the study of the Cro and cI repressor proteins from the bacteriophages λ and 434. The amino acid side-chains of the recognition helix in the monomeric unit read the patten of hydrogen-bond acceptors, hydrogen-bond donors, and van der Waals contacts presented in the major groove for one-half of the operator. However, both monomers of the dimeric repressor must align properly with each half-site in the operator sequence. The results thus far suggest a rather passive role for DNA—in the absence of protein the DNA has a regular conformation. The specificity of recognition seems to rely more on the unique amino acid sequence and structural fold of the protein. As will be shown below, this may not be true for all DNA-protein recognition systems, even with proteins containing the helix-turn-helix DNA-binding structural motif.

3. Protein-DNA Interactions in the Trp repressor-operator complex

The main mechanism for the transcriptional control of L-tryptophan levels in *E. coli* is the repression of tryptophan biosynthesis by end-product inhibition. In the absence of tryptophan, the Trp repressor is unable to bind DNA, no repression of transcription occurs for the *trp* operon, and enzymes required for the biosynthesis of tryptophan are made. The binding of tryptophan to the Trp repressor increases the affinity of the repressor for operator DNA, and down-regulates transcription in the *trp* biosynthesis operon. The crystal structure determination (Shevitz et al., 1985) of the Trp repressor with its corepressor tryptophan reveals the helix-turn-helix DNA-binding substructure, as suggested by amino acid homologies (Figure I-10). In the absence of tryptophan, the relative positions of the helices of the motif change in that they no longer match the shape of B DNA. Ligand-induced conformation changes may be a common mechanism in the regulation of DNA binding by allosteric effector molecules.

In the structure of the Trp repressor bound to an 18 base-pair operator consensus sequence, the protein is relatively unchanged from its unbound state (Otwinowski et al., 1988). The DNA shows only Watson-Crick hydrogen-bonding and is bent towards the ends as a cumulative result of small local helix distortions. Such bending allows a good fit of protein and DNA, with 2900 $Å^2$ of contact surface inaccessible to bulk solvent. There is extensive hydrogen bonding between protein side-chains and the DNA, and a number of water-mediated protein-DNA bonds. What is unusual, or novel, about the complex is that there are no direct hydrogen-bonds, or van der Waals contacts, between the amino acid side-chains of the recognition helix and the base-pairs in the major groove of the DNA.

It was argued (Sigler et al., 1988) that the Trp repressor and the operator DNA recognize each other by an indirect mechanism with a combination of protein-water-DNA hydrogen bonds and by sequence-dependent conformational effects of the DNA. The water molecules were considered an extension of the hydrogen-bonding surface of the protein (or DNA) and could be regarded as a variation on the mechanism used by the DNA-binding proteins discussed earlier. The selectivity by the indirect mechanism is somewhat lower than that achieved in some protein-DNA complexes using direct contacts between protein and DNA. The *in vitro* preference of Trp repressor for specific operator over a random non-operator sequence is 10^4, whereas the Cro repressor has a 10^6-fold preference (Kim et al., 1987) and the Lac repressor has a 10^8-fold preference for specific over non-specific DNA (reviewed by Travers, 1989).

However, it is difficult from the structure of the complex to explain key biochemical studies on mutant Trp repressors (reviewed in Brennan & Matthews, 1989a) and, since the

changes in the DNA structure are not large from the conformation of classical B DNA, it is difficult to see how they would be sufficient for sequence-dependent recognition of the Trp repressor. The crystallization conditions might have favoured formation of a non-specific complex. This is more consistent with the structural features of the Trp protein-operator DNA complex. The N-terminal arms of the protein (residues 2-14), which enhance both non-specific and specific DNA binding, are disordered in the structure. Notwithstanding these doubts expressed over the specificity of the observed complex, it seems likely that the conformation of the DNA has a role for distinguishing the specific operator DNA sequence from other sites on DNA.

4. Studies of the Lac repressor and Lac repressor-DNA interactions

The expression of the gene products required for lactose utilization in *E. coli* is regulated by binding of the Lac repressor to an operator site in the *lac* operon. The N-terminal portion of the sequence displays homology with other DNA binding proteins (Figure I-9), and is therefore expected to contain a helix-turn-helix DNA-binding unit. A model for the Lac repressor-operator complex has been proposed by Matthews et al. (1982) analogous with the model for the Cro-O_R3 complex. Proteolytic cleavage of the intact repressor results in an N-terminal domain, or headpiece, (residues 1 to 51, or 1 to 56) that binds DNA, making the same contacts to the DNA as the full repressor. In solution, the Lac repressor headpiece is monomeric. Two headpieces can bind to the pseudo-two-fold symmetric 21-25 base-pair *lac* operator, one in each half-site (Nick et al., 1982; Scheek et al., 1983). Because the N-terminal domain is small, and the biochemistry and genetics of the *lac* operon is well characterized, the Lac repressor headpiece is an ideal candidate for studies of its structure and its interaction with DNA. Kaptein and co-workers have employed the novel procedure of using nuclear magnetic resonance techniques for structure determination of both the headpiece and its complex with DNA (Kaptein et al., 1985; Lamerichs et al., 1989).

Lac repressor headpiece (Residues 1-51) is all helical, with the relative orientation of helices I and II similar to the of the helix-turn-helix motif (α2-α3 helices) of λ Cro repressor. The headpiece has a central hydrophobic core, stabilizing the structure of the monomer in solution. No clear indication for a possible dimer interface region was seen in the structure that could give information on the relative orientation of two headpiece molecules when bound to DNA.

From analogy to the other helix-turn-helix proteins, it had been proposed that the recognition helix is positioned in the major groove of the DNA with the N-terminal end

near the outer part of the operator (Matthews et al., 1982). For example, in the Cro-operator complex tyrosine 26, at the N-terminus of the α3 recognition helix contacts bases at +1 of the operator sequence, whereas His 35, in the middle of the helix, contacts phosphates +4 or +5, and Arg 38, in the loop after the recognition helix contacts bases -6 and -7, close to the center of the operator (Figure I-9). This binding mode is also seen in the crystal structure of the λ repressor-operator complex (Figures I-14 and I-15). Positioning the Lac headpiece in a similar orientation would place Tyr 7 close to base-pairs 3 and 4, and His 29 to make contact near the center of the operator at base-pairs 9 or 10 (Figure I-16).

Intermolecular NOE crosspeaks between the Lac repressor headpiece and DNA tell a quite different story. NMR solution structure studies were carried out with a 56 residue Lac headpiece-half operator complex and with the Lac headpiece-full operator (22 bp) complex (Figure I-16; Boelens et al., 1987; Lamerichs et al., 1989). The 3,5 protons of Tyr 7 make a close contact to the H8 proton of G10, near the center of the operator. Additionally, a NOE crosspeak is observed between the C2 proton of His 29 and the methyl protons of both T3 and T4. The orientation of the recognition helix is reversed in the major groove as compared to the recognition helices of the aforementioned repressors, with the N-terminal end of the helix near to the center of the operator's two-fold symmetry.

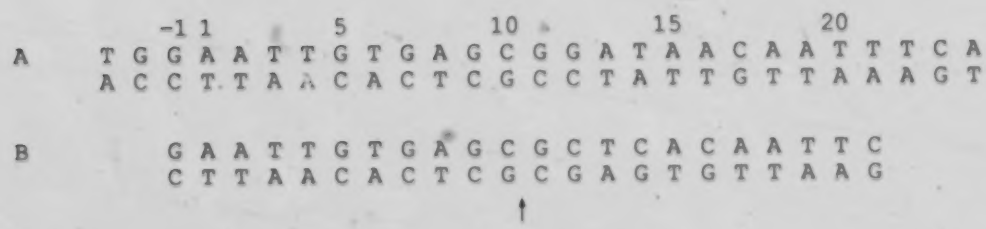

-1 1 5 10 15 20
A
T G G A A T T G T G A G C G G A T A A C A A T T T C A
A C C T T A A C A C T C G C C T A T T G T T A A A G T
B
G A A T T G T G A G C G C T C A C A A T T C
C T T A A C A C T C G C G A G T G T T A A G

Figure I-16 Operator DNA sequence of the *lac* operon

The DNA sequence of the operator site in the *lac* operon is shown in the upper sequence (A).The numbers are used to denote base-pair position in the operator (from Lamerichs et al., 1989). The half-operator used to form the protein-DNA complex in the initial study covered base-pairs -2 to 12. In B, the tight-binding fully symmetric *lac* operator is shown with an arrow indicating the center of the operator.

The reverse orientation of the recognition helix is confirmed by other observed intermolecular NOE crosspeaks, and is supported by genetic experiments of the intact Lac repressor, and of the related Gal repressor (Lamerichs et al., 1989; Lehming et al., 1987).

The helix-turn-helix substructure is often used for DNA-binding, but not necessarily in the same manner. In initial search of the Brookhaven Protein Data Bank failed to find any similar two-helical unit (Takeda et al., 1983) in a protein that does not specifically bind duplex DNA. However, subsequently backbone structures that correspond well to the helix-turn-helix have been located in cytochrome c peroxidase and in the L7/L12 ribosomal proteins (Brennan & Matthews, 1989a). Although the rms deviation (about 1.4 Å) is not as good as is observed with other helix-turn-helix DNA binding proteins (less than 1 Å), the structures found suggest that the helix-turn-helix is a structural motif of proteins, but not necessarily one exclusively used for sequence-specific recognition of duplex DNA.

In summary, the recognition helix of the helix-turn-helix is a convenient way to present the amino acid side chains of the protein so that they can make a reasonable number of contacts within the major groove of DNA. This motif occurs as a pair in a dimeric protein, and must align with the pair of affinity sites presents in each half of the operator DNA. Different proteins use the motifs in different ways. The helix-turn-helix motif is not the only structural unit capable of interacting and binding strongly to a specific site on DNA. Other DNA-binding proteins, such as the EcoRI endonuclease and the Zn finger proteins use quite different mechanisms.

5. The DNA-Eco RI endonuclease recognition complex

Some bacteria have a defence against the invasion of foreign DNA, such as the DNA of viruses. Certain strains of *E. coli* produce enzymes with an endonuclease activity that restrict the action of some bacteriophages. One of these enzymes, called EcoRI restriction endonuclease recognizes a two-fold symmetric six base pair sequence (Figure I-17) and cleaves both strands between the guanine and adenine. The host bacterium protects its own DNA by methylation of the central adenines of both strands at the exocyclic N6 amino group. When either one or both groups are methylated, EcoRI no longer cleaves the DNA. Therefore, not only does the endonuclease distinguish this hexanucleotide sequence from all others, it also discriminates between different methylation states of the same nucleotide. A lack of fidelity on the part of the host would result in cleavage of its own DNA. The DNA repressors so far discussed had DNA binding sites extending from 14 to over 20 base-pairs. This recognition site is much more compact, so that the protein must use a different mechanism for DNA recognition.

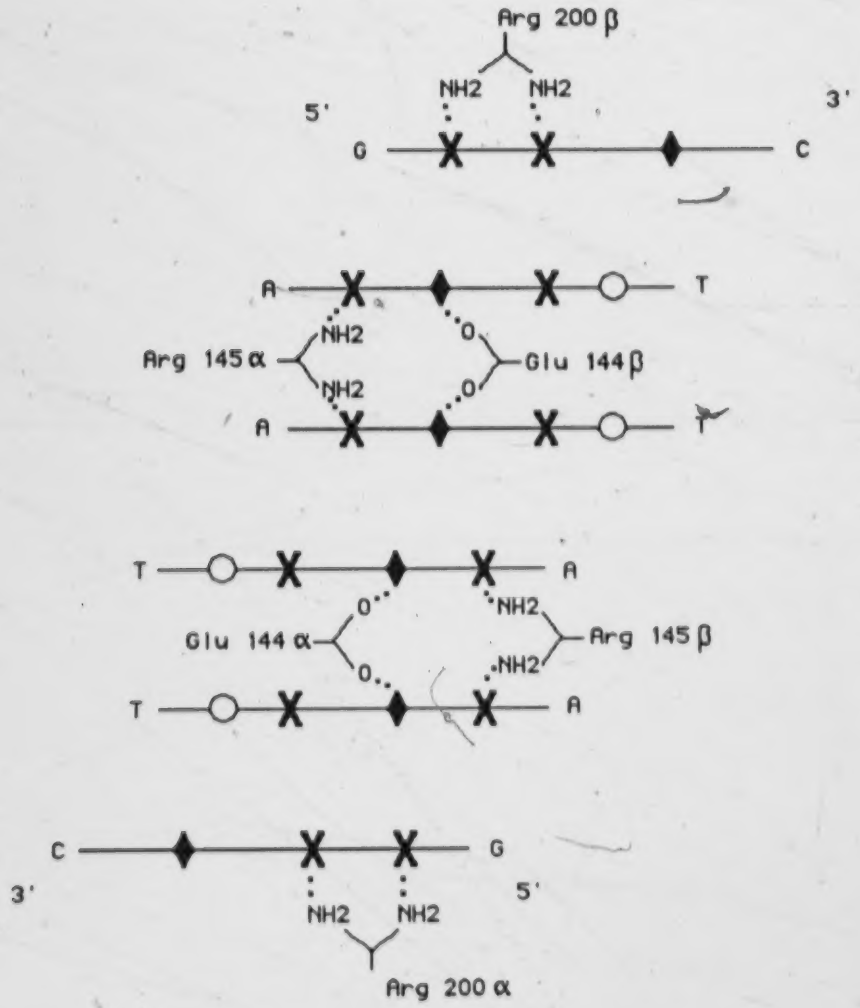


Figure I-17 Sequence specific interactions between DNA and EcoRI endonuclease
This figure shows a schematic representation of the sequence-specific interactions between the protein EcoRI endonuclease and its DNA site based on the crystal structure (McClarin et al., 1986). Nomenclature and symbols are given on Figure I-9.

The endonuclease has identical subunits of 31kD and requires Mg^{2+} for phosphodiester bond hydrolysis. In the absence of Mg^{2+}, the protein forms a stable sequence-specific complex with DNA containing the cognate hexanucleotide sequence, with a dissociation constant (in solution) on the order of 10^{-11} M^{-1}. The 3 Å resolution crystal structure of the complex between EcoRI endonuclease and DNA provides yet another detailed example of the structural basis of sequence-specific DNA-protein interactions.

The complex has a two-fold symmetry, with the two fold axis coincident with the center of symmetry of the DNA sequence. Each subunit of the protein is organized into an α/β domain with α helices, a five strand β sheet, and an extension that wraps around the DNA. As in Cro repressor, the β sheet forms the foundation for the α helices responsible for sequence-specific recognition. The outer two guanines are recognized with bidentate hydrogen bonds by an arginine side-chain of an α helix from each subunit (Figure I-17). The central AATT is recognized by separate α-helices, one from each subunit. In each subunit a glumate and arginine form bidentate hydrogen bonds that bridge adjacent adenine residues. These twelve hydrogen-bonds are sufficient for discrimination of the GAATTC sequence because any base substitution would require rupture of at least one of the hydrogen bonds.

Binding of the protein induces localized distortions in the DNA. They are not seen in the crystal structure for the unbound DNA of nearly identical sequence $[d(CGCGAATTCGCG)]_2$ (Dickerson & Drew, 1981) and are termed neo-kinks. The type I neo-kink introduces a net unwinding of 25° into the DNA near the center of the sequence, and aligns the adjacent adenine residues to reflect the geometry necessary for the glutamate and arginines to bridge across the purines. Other neo-kinks occur at the edges of the six base-pair recognition element (Figure I-17).

The recognition geometry is stabilized by interactions between amino acid side-chains in the protein, and includes electrostatic interactions between oppositely charged pairs Glu_{144}-Arg_{145} and Glu_{144}-Arg_{200}. This is suggested (McClarin et al., 1986) to sharpen the discrimination between cognate and noncognate sites in that the DNA interaction energy is not a simple additive sum over the individual interactions (as with Cro, Figure I-10). Formation of some correct protein-base interactions facilitates the formation of additional correct interactions, whereas incorrect interactions with non-cognate DNA sites would have an inhibitory effect.

The very high specificity of EcoRI endonuclease derives from a series of sequence-specific steps including DNA-binding to the correct site, and then allosteric activation to induce the capability to hydrolyze the phosphodiester bond. Errors are corrected at each step by dissociation of non-cognate DNA-protein complexes, without activation of the

enzyme resulting in a very low cleavage rate at non-cognate sites. McClarin et al. (1986) suggest that an enzyme that covalently modifies DNA is intrinsically capable of achieving a much higher level of sequence discrimination than is a simple binding protein.

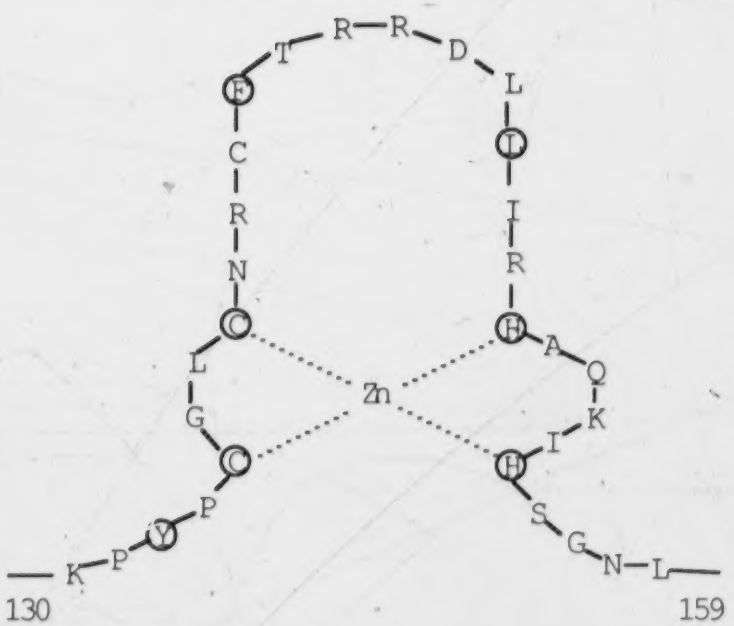
The EcoRI recognition motif is clearly different from that of the aforementioned helix-turn-helix DNA-binding proteins. This difference may be due to the high specificity demanded of a restriction enzyme, or to the highly concentrated nature of the EcoRI recognition site. Nonetheless, the difference shows that the helix-turn-helix is not a universal DNA recognition element.

6. Zinc finger DNA-binding proteins

Recently, sequence analysis of TFIIIa, a transcription factor protein required for correct initiation of 5S RNA genes by RNA polymerase III, has led to the proposal of a distinct DNA-binding motif, the zinc finger (Miller et al., 1985). TFIIIa binds 7 to 11 zinc atoms per molecule and contains nine repeating units of approximately 30 residues. Each unit is composed of two invariant pairs of cysteines and histidines with the consensus sequence: (Tyr-Phe)-X-Cys-X_4-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His-X_{2-6}, where X is not a conserved amino acid, and the cysteines and histidines are thought provide the ligands to coordinate Zn^{2+} ions (Figure I-18). Subsequently, similar units have been found in other transcription factors, and in other nucleic acid binding proteins (reviewed in Klug & Rhodes, 1987; Struhl, 1989).

Párraga et al. (1988) have used a variety of spectroscopic techniques to characterize a single zinc finger domain (residues 130-159) from yeast transcription factor, ADR1a (Figure I-18). ADR1 is a positive transcription regulator of the glucose-repressible alcohol dehydrogenase gene in *Saccharomyces cerevisiae*. Sequence analysis of its 1323 amino acid residues reveals two Zn finger domains between residues 100 and 160. Mutations producing null alleles are clustered in this region and an ADR1(17-229)-β galactosidase fusion protein footprints activator sequence DNA.

The single domain Zn finger also binds Co^{2+} with absorbance maxima characteristic for tetrahedral coordination of the metal ion and thio-containing ligands. The Co^{2+} absorption bond disappears when a slight molar excess of Zn^{2+} is added, suggesting that Zn^{2+} occupies the same metal pocket and ligand geometry as the Co^{2+} ion. Circular dichroism (CD) spectra reveals zinc promoted α-helicity for approximately 10 residues of the 30 residue domain. Two-dimensional NMR nuclear Overhauser effect spectroscopy (NOESY) shows crosspeaks between sequential amide protons for residues 135 to 137, and residues



ADR1a

Figure I-18 Zinc finger amino acid sequence

Shown is the amino acid sequence of one of the Zn finger DNA-binding domains, ADR1a, of the yeast transcription activator, ADR1. Amino acids residues conserved between different Zn fingers are circled. Histidine and cysteine residues coordinate the central Zn^{2+} ion (from Párraga et al., 1988).

146 to 153. A model has been proposed for which the Zn is tetrahedrally coordinated, with residues 135 to 137 forming a short turn, followed by a loop, and then another three turns of an α helix for residues 146-153 (Figure I-18; Párraga et al., 1988). Although this single 30 residue domain is stably folded, it does not bind to a specific sequence of DNA. This suggests that multiple fingers are required to contribute the correct number of base contacts for sequence-specific binding, or that the single domain requires a larger protein context to obtain a proper geometry for specific DNA-binding. The correspondence of the α-helix seen for this single domain with the mutations that interfere with DNA-binding suggest that the α-helix could take part in binding to the major groove in manner not entirely dissimilar to that of the helix-turn-helix proteins. However, the exact mode of interaction of Zn fingers with DNA is still poorly understood (Párraga et al., 1989).

D. The structural basis of DNA-protein recognition

1. General features of DNA-protein recognition

A number of different ways in which recognition of protein and DNA occur have been presented. Other sequence specific DNA recognition systems exist, such as those involved in replication (e. g. the Klenow fragment of DNA polymerase structure is known, Ollis et al., 1985) and other DNA-recognition motifs have been studied. For example, the β sheet residues of the P22 Arc repressor and the *E.coli met* repressor are responsible for the specificity of operator DNA binding (recently reviewed in Bowie & Sauer, 1989; Rafferty et al., 1989; see also Brennan & Matthews, 1989a; Struhl, 1988). Undoubtedly, new methods of DNA-protein recognition are to follow. In the recognition between DNA and proteins that regulate transcription, the DNA plays a mainly passive role by presenting a pattern of hydrogen bond and van der Waals contacts to the protein and not by dramatically altering its conformation in a unique fashion at the specific site of binding.

In other protein-DNA systems, the conformation of the DNA plays an important role for weak site-specific recognition by proteins. For example, the binding of histones to DNA to form nucleosomes is DNA sequence dependent (Richmond et al., 1984; Drew & Travers, 1985; Travers, 1989). Nucleosome positioning corresponds to bending of DNA around the histone octamer core, and as such the conformation of the DNA determines the protein binding site. Certain sequences of DNA bend even in the absence of protein, and the structural basis of DNA bending is beginning to be understood (for examples see Travers, 1989; Nadeau & Caruthers, 1989).

In summary, the DNA binding domain has two important roles in transcriptional regulation: (1) it brings the protein to the DNA such that they can interact to bring about a functional transcriptional initiation complex; and (2), high specificity of protein-DNA interaction provides a major mechanism by which genes are differentially expressed. In regard to this function, the distinct DNA-binding sub-structures are often equivalent.

Amino acid side-chains are held in position by the rest of the protein to read the pattern of possible hydrogen-bonds and van der Waals contacts presented by the DNA. Multiple but roughly equivalent binding sub-sites are contained within the DNA operator that are contacted by multiple sub-structures on the protein. In the case of helix-turn-helix repressor proteins and the EcoRI endonuclease protein-DNA recognition systems, DNA, protein and DNA-protein complex are all approximately two-fold symmetric. The half sites of the protein and DNA must align properly for strong interaction between protein and DNA. Many of the DNA-binding proteins are small and are good candidates for structural studies in solution using nuclear magnetic resonance techniques.

2. Prospects for altering DNA-protein recognition

The introduction of molecules to regulate transcription at specific locations on DNA into the cell is a means to control certain enzymatic activities associated with disease. One example would be to suppress oncogene product levels in tumors, leading to more effective ways of dealing with cancer. Alternatively, transcription of the appropriate genes is required to correct inborn problems of metabolism caused by the lack of sufficient quantity of a metabolic enzyme. The sequence-specific repressors include molecules administered as drugs to a patient, such as some chemotherapeutic agents. The minor groove DNA-binding drugs, and even unnatural DNA molecules (such as α-oligonucleotides, which have an inversion of chirality about the C1' atom) show great progress in this regard. Progress continues for use as therapeutic agents, especially with regard to delivery to the site of action within the body, reducing toxic side-effects, and improving sequence specificity.

Gene therapy techniques are now becoming available that introduce foreign DNA into a mammalian system and provide a means to produce proteins within the cell (for examples, see the recent report (1989) on *Genes in Medicine* under the News and Comment section in *Science 242*, 746-751). We have already seen that DNA-binding proteins efficiently repress transcription in many biological processes. Can they be altered to bind to any desired sequence to change the level of transcription?

A full appreciation of the complexity and individuality of each of the aforementioned complexes seems to be, at first, discouraging for anyone hoping to find simple answers to

the recognition problem. However, there are clear examples of protein engineering having been successfully applied to repressors of transcription. Here, I will outline two attempts to change the specificity of helix-turn-helix DNA-binding proteins.

Wharton & Ptashne (1985) have replaced four amino acids on the solvent-exposed surface of the recognition helix of 434 repressor with the corresponding amino acids from the recognition helix of a different repressor, that encoded by the *Salmonella* bacteriophage P22. The P22 repressor is proposed to be a helix-turn-helix protein (Figure I-10) and the phage 22 is another close relative of the phages λ and 434. Phage 434 repressor binds to P22 operator DNA only non-specifically and *vice versa*. The re-designed 434 repressor, with the P22 recognition amino acids, has the binding specificity of P22 repressor, as measured *in vivo* and *in vitro*. The specificity changed, for the most part, from that of the 434 sequence to that of the P22 sequence. The re-designed 434 repressor can no longer bind specifically to 434 operator. These results suggest that a simple set of rules could be devised to design a protein to bind to a desired DNA sequence using the foundation of a given DNA-binding protein.

The Cro repressor has also been re-designed to recognize a new operator (Caruthers et al., 1987). The affinity of wild type and site-specific mutants of Cro repressor was measured for DNA containing an O_R1 operator DNA sequence. Gln 27 of Cro is proposed to make a hydrogen bond to the N7 and exocyclic N6 of the adenine at position 2 in each half of the operator DNA (Figure I-9). If the A:T base-pair at this position is replaced by a T:A base-pair (to create a O_R1* operator), a greater than 40 fold loss in affinity occurs for wild type Cro (Table I-1). Substituting the hydrophobic amino acids Cys, Leu, or Val for Gln 27 lowers the affinity of Cro for O_R1. However, these mutant repressors all have an affinity to O_R1* approximately equal to the affinity of wild type protein for wild type O_R1 operator DNA. Replacement of thymine 2 by uracil results in loss of affinity of the mutant Cro repressors for O_R1*. These results suggest that substituting a hydrophobic amino acid at Gln 27 does not significantly alter the three-dimensional structure of Cro and that the hydrophobic interaction proposed to be made to the thymine methyl group is approximately equivalent to the hydrogen bonding interaction to adenine N7 in the wild-type Cro - wild type operator system. Presumably these results would also hold true for binding to the O_R3 and substituted O_R3 operator DNA sequences. These re-engineered proteins recognize a different DNA sequence and represent a first step towards the objective of being able to construct a DNA-binding protein to interact with any desired DNA sequence.

Table I-1 Interaction of Cro repressor with operator DNA

	K_d (nM)		
	O_R1	O_R1*	O_R1*-uracil
Protein			
Gln 27 (wild type)	2.5	>100	>100
Cys 27	20.0	2.0	35.0
Leu 27	4.5	1.5	11.0
Val 27	6.0	1.7	16.0

Equilibrium dissociation binding constants for the binding of the Cro repressor to operator DNA were measured by a gel retardation assay. The O_R1* sequence has a transversion of the AT base-pair at position 2 in each half of the operator (using the nomenclature of Figure I-9). In O_R1*-uracil the thymines at position 2 are replaced by uracil. The wild type Cro repressor protein has Gln at position 27 in its amino acid sequence. Site-specific mutants contain different amino acids at position 27 (adapted from Caruthers et al., 1987).

3. Outline of the thesis

The structural basis of protein-DNA recognition has been reviewed in the introduction to this thesis. Particular attention was paid to the family of proteins containing a helix-turn-helix DNA-binding motif, which includes the Cro repressor protein from bacteriophage λ. Considerable attention in the literature has focussed on protein being the active participant in the recognition process. Here, emphasis was placed on the structural role of DNA. The groundwork for the determination of nucleic acid structure in solution using nuclear magnetic resonance techniques is illustrated in Chapter II. Mathematical procedures for distance determination, NOE data assessment, and NOE-based refinement are worked out with a DNA decamer comprising the major sequence-specific binding site for the Cro repressor protein. In Chapter III, these NMR methods are used to determine the solution conformation of two self-complementary DNA octamers. The DNA duplexes are of alternating purine-pyrimidine sequence and demonstrate the diversity of DNA structure. In Chapter IV, we return to the more difficult DNA decamer, and determine its structure in solution. The DNA decamer duplex comprises the left ten base-pairs of the O_R3 operator sequence from phage λ. The interaction of this half-operator with wild-type and mutant Cro repressor proteins is detailed in Chapter V. The formation of the complex involving the left ten base-pair half-operator and the native Cro repressor is studied in detail. Changes induced in the conformation of the DNA by binding to protein are observed. The O_R3 operator sequence is pseudo-symmetric, with its symmetry axis co-existing with the two-fold symmetry axis of the Cro protein. By characterizing the binding strengths and stoichiometries of various half- and full-operator DNA sequences, the role of symmetry in protein-DNA interaction is highlighted (Chapter VI). In the last chapter, implications of these studies into the DNA-Cro repressor system from bacteriophage λ are discussed in terms of general protein-DNA recognition.

E. References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899-907.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754-758.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504-1509.

Bax, A. (1989) *Ann. Rev. Biochem. 58*, 223-256.

Bax, A. (1982) *Two-dimensional Nuclear Magnetic Resonance* Reidel, Boston.

Boelens, R., Scheek, R. M., van Boom, J. H., & Kaptein, R. (1986) *J. Mol. Biol. 193*, 213-216.

Boschelli, F., Arndt. K., Nick, H., Zhang, Q., Lu, P., & Takeda, Y. (1982) *J. Mol. Biol. 162*, 251-266.

Bowie, J. U., & Sauer, R. T. (1989) *Biochemistry 28*, 7139-7143.

Brennan, R. G., & Matthews, B. W. (1989a) *Trends Biochem. Sci. 14*, 286-290.

Brennan, R. G., & Matthews, B. W. (1989b) *J. Biol. Chem. 264*, 1903-1906.

Caruthers, M. H., Barone, A. D., Beltman, J., Bracco, L. P., Dodds, D. R., Dubendorff, J. W., Eisenbeis, S. J., Gayle, R. B., Prosser, K., Rosendahl, M. S., Sutton, J., & Tang, J.-Y. (1986) *in Protein Structure, Folding, and Design* (Oxender, D. L., ed.) pp. 221-228, Liss, New York.

Caruthers, M. H., Gottlieb, P., Bracco, L., & Cummins, L. (1987) in *Protein Structure, Folding, and Design 2* (Oxender, D. L., ed.) pp. 9-24, Liss, New York.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761-786.

Drew, H. R., & Travers, A. A. (1985) *J. Mol. Biol. 186*, 773-790.

Eisenbeis, S. J., Nasoff, M. S., Noble, S. A., Bracco, L. P., Dodds, D. R., & Caruthers, M. H. (1985) *Proc. Natl. Acad. Sci. 82*, 1084-1088.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford.

Folkmanis, A., Takeda, Y., Simuth, J., Gussin, G., & Echols, H. (1976) *Proc. Natl. Acad. Sci. 73*, 2249-2253.

Friedman, D. I., Olson, E. R., Georgopoulos, C., Tilly, K., Herskowitz, I., & Banuett, F. (1984) *Microbiological Reviews 48*, 299-325.

Hendrix, R. W., Roberts, J. W., Stahl, F. W., Weiberg, R. A. (ed.)(1983) *Lambda II*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Hochschild, A., Irwin, N., & Ptashne, M. (1985) *Cell 32*, 319-325.

Hsiang, M. W., Cole, R. D., Takeda, Y., & Echols, H. (1977) *Nature 270*, 275-277.

Jacob, F., & Monod, J. (1961) *J. Mol. Biol. 3*, 318-356.

Jarema, M. C., Lu, P., & Miller, J. H. (1981) *Proc. Natl. Acad. Sci.78*, 2207-2711.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893-899.

Kaptein, R., Boelens, R., Scheek, R. M., & van Gunsteren, W. F. (1988) *Biochemistry 27*, 5380-5395.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179-182.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149-158.

Kirpichnikov, M. P., Yartzev, A. P., Minchenkova, L. E., Chernov, B. K., & Ivanov, V. I. (1985) *J. Biomol. Struct. Dynam. 3*, 529-536.

Klug, A., & Rhodes, D. (1987) *Trends Biochem. Sci. 12*, 464-469.

Koudelka, G. B., Harbury, P., Harrison, S. C., & Ptashne, M. (1988) *Proc. Natl. Acad. Sci. 85*, 4633-4637.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1988) *Nature 326*, 886-888.

Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Rüterjans, H. (1989) *Biochemistry 28*, 2985-2991.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., & Takeda, Y. (1982) *Proc. Natl. Acad. Sci. 79*, 1428-1432.

McClarin, J. A., Frederick, C., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526-1541.

Metzler, W. J., & Lu, P. (1989) *J. Mol. Biol. 205*, 149-164.

Metzler, W. J., Arndt, K., Tecza, E., Wasilewski, J., & Lu, P. (1985) *Biochemistry 24*, 1418-1424.

Miller, J., McLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609-1614.

Mondragón, A., Subbiah, S., Almo, S. C., Drottar, M., & Harrison, S. C. (1989) *J. Mol. Biol. 205*, 189-200.

Mondragón, A., Wolberger, C., & Harrison, S. C. (1989) *J. Mol. Biol. 205*, 179-188.

Morris, G. A. (1986) *Magn. Reson. Chem. 24*, 371-403.

Nadeau, J. G., & Caruthers, D. M. (1989) *Proc. Natl. Acad. Sci. 86*, 2622-2626.

Nick, H., Arndt, K., Boschelli, F., Jarema, M. C., Lillis, M., Sadler, J., Caruthers, M., & Lu, P. (1982) *Proc. Natl. Acad. Sci. 79*, 218-222.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718-723.

Ohlendorf, D. H., & Matthews, B. W. (1983) *Ann. Rev. Biophys. Bioeng. 12*, 259-284.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985) *Nature 313*, 762-766.

Otwinowski, Z., Schevitz, R.W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321-329.

Pabo, C. O., & Lewis, M. (1982) *Nature 298*, 443-447.

Pabo, C. O., & Sauer, R. T. (1984) *Ann. Rev. Biochem. 53*, 293-321.

Pakula, A. A., Young, V. B., & Sauer, R. T. (1986) *Proc. Natl. Acad. Sci. 83*, 8829-8833.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Beyer, B. J., Pabo, C. O., Roberts, T. M., & Sauer, R. T. (1980) *Cell 19*, 1-11.

Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489-1492.

Párraga, G., Young, L., & Klevit, R. E. (1989) *Trends Biochem. Sci. 14*, 398.

Rafferty, J. B., Sommers, W. S., Saint-Giron, I., & Phillips, S. E. V. (1989) *Nature 341*, 705-710.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984) *Nature 311*, 532-537.

Roberts, T. M., Shimatake, H., Brady, C., & Rosenberg, M. (1977) *Nature 270*, 274-275.

Sarai, A., & Takeda, Y. (1987) in *Protein Structure, Folding, and Design 2* (Oxender, D. L., ed.) pp. 57-64, Liss, New York.

Sarai, A., & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. 86*, 6513-6517.

Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom, J. H., Kaptein, R., Rüterjans, H., & Beyreuther, K. (1983) *Biochemistry 22*, 228-235.

Shea, M. A., & Ackers, G. K. (1985) *J. Mol. Biol. 181*, 211-230.

Shevitz, R. W., Otwinowski, Z., Joachmiak, A., Lawson, C. L., & Sigler, P. B. (1985) *Nature 317*, 782-786.

Shirakawa, M., Lee, S. J., Akutsu, H., Kyogoku, Y, Kitano, K., Shin, M., Ohtsuka, E., & Itekura, M. (1985) *FEBS Letters 181*, 286-290.

Struhl, K. (1989) *Trends Biochem. Sci. 14*, 137-140.

Takeda, Y., Folkmanis, A., & Echols, H. (1977) *J. Biol. Chem. 252*, 6177-6183.

Takeda, Y., Kim, J. G., Gaday, C. G., Steers, E., Jr., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1986) *J. Biol. Chem. 261*, 8608-8616.

Takeda, Y., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1983) *Science 221*, 1020-1026.

Takeda, Y., Sarai, A., & Riviera, V. M. (1989) *Proc. Natl. Acad. Sci. 86*, 439-443.

Travers, A. A. (1989) *Ann. Rev. Biochem. 58*, 427-452.

van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1-38.

Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985) *Biochemistry 24*, 4553-4562.

Wemmer, D. E., Chou, S.-H., & Reid, B. R. (1984) *J. Mol. Biol. 180*, 41-60.

Wharton, R. P., & Ptashne, M. (1985) *Nature 316*, 601-605.

Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature 335*, 789-795.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Chapter II

Distance Measurement and Structure Refinement with NOE data[1]

[1]A version of this chapter has been accepted for publication: Baleja, J. D., Moult, J., & Sykes, B. D. (1990) *J. Magn. Reson. 87*, 000-000. Reprinted, with permission, from Baleja, J. D., Moult, J., & Sykes, B. D. (1990) *J. Magn. Reson. 87*, 000-000. Copyright (1990) Academic Press.

A. Introduction

In the first chapter of this thesis, a procedure was briefly described for the determination of the structure of proteins in solution using nuclear magnetic resonance techniques. In Chapters III and IV, the solution structures are determined for three nucleic acid molecules. In this chapter, new methods are presented for estimating distances between protons and for assessing and refining structures with NOE data. These methods are developed using the observed and calculated NOE intensities for a DNA duplex, although they would apply equally well to other biological macromolecules, such as proteins and carbohydrates.

The two-dimensional ^{1}H NMR nuclear Overhauser experiment is a powerful method to determine the spatial proximity of protons within a macromolecule. Computation of molecular conformation can be based on a set of distances derived from NOESY cross-peaks (Noggle & Shirmer, 1971; Havel & Wüthrich, 1985; Kaptein et al., 1985; Wüthrich, 1986). Often a cross-peak intensity is taken as inversely proportional to the sixth power of the distance between two spins[2]. Distances so derived are only approximate since the cross-peak intensity due to direct cross-relaxation between spins i and j is modified by additional cross-relaxation with any spin k, especially if spin k exists such that $r_{ik} < r_{ij}$ or $r_{jk} < r_{ij}$. Accuracy is improved with a proper correction for spin diffusion and since longer distances are affected more than shorter ones, the range of distances can be extended to better define the conformational space covered by the molecule. Spin diffusion effects can be eliminated by two methods that involve the calculation of the NOE from a given structure (Macura & Ernst, 1980; Bodenhausen & Ernst, 1982). Distances may be corrected by back-transformation of a combined experimental and calculated NOE matrix (Olejniczak et al., 1986; Boelens et al., 1989). Alternatively, calculated NOE intensities may be compared directly to observed intensities and the initial structure can be adjusted in an iterative manner so as to minimize the difference between the two sets of NOEs (Suzuki et al., 1986; Lefèvre et al., 1987; Borgias & James, 1988). However, the computational requirements for a complete NOE calculation are cumbersome and therefore approximate distances have been most often used to determine molecular structure in solution by NMR (Wüthrich, 1986; Lefèvre et al., 1987).

The first section of Chapter II evaluates three procedures for the best method to determine interproton distances from NOE data. Using the method of highest performance,

---

[2]An example of a ^{1}H NMR NOESY spectrum is given in Chapter III (Figure III-2). The closer that two protons are, the stronger the resulting interaction and corresponding cross-peak between them.

interproton distances are then estimated for a DNA decamer, d(TCTATCACCG) ·d(CGGTGATAGA), which comprises the left ten base-pairs (L10) of the bacteriophage λ O_R3 operator. An initial structure is derived by using restrained molecular dynamics calculations (Nilges et al., 1987) with the basis set of approximate distances, and with a set of allowed torsion angles.

The second part of the chapter describes the incorporation of NOE-based refinement into NMR structure determination. The direct comparison between calculated and observed NOE intensities is used as an effective energy potential in a restrained molecular mechanics calculation. The calculated NOEs are evaluated using a full matrix analysis procedure, and include the effects of spin diffusion and differential motion within the molecule. Energy minimization results in structures which are in agreement with the experimental data, as evaluated by an R factor analogous to the standard crystallographic R factor (Baleja & Sykes, 1988; Gupta et al., 1988).

B. Theory

Measurement of the ^1H NMR nuclear Overhauser effect probes the extent of the relaxation pathways between protons. Relaxation processes describe the return of the magnetization to equilibrium after a perturbation has been applied to populations in the nuclear spin system. The theory of NMR relaxation has been extensively characterized (Abragam, 1961; Solomon, 1955; Slichter, 1978; Harris, 1983; Ernst et al., 1987) and is only briefly described here.

Homonuclear dipolar relaxation for a molecule tumbling isotropically in solution can be described by (Solomon, 1955; Abragam, 1961):

$$\frac{d\mathbf{A}}{d\tau_m} = -\mathbf{R} \cdot \mathbf{A} \qquad (1)$$

for which **A** is the matrix of magnetizations[3] and **R** is the cross-relaxation matrix with diagonal elements:

$$R_{ii} = Q \sum_{(i \neq j)} \frac{1}{r_{ij}^6} \left[J_0(\omega) + 3 J_1(\omega) + 6 J_2(\omega) \right] \qquad (2)$$

and off-diagonal elements:

---

[3]The difference between the component of the magnetization along the z direction and its equilibrium value $(M_z - M_0)$.

$$R_{ij} = \frac{Q}{r_{ij}^6} [\, 6 J_2(\omega) - J_0(\omega) \,] \qquad (3)$$

where Q is equal to $0.1\gamma^4\hbar^2$, and the spectral densities, $J_n(\omega)$, take the form:

$$J_n(\omega) = \frac{\tau_{ij}}{1+(n\omega_0\tau_{ij})^2} \qquad (4)$$

with τ_{ij}, the rotational correlation time of the interproton vector between i and j, and ω_0, the Larmor frequency. Cross correlation terms between different pair-wise dipolar interactions (Bull, 1987) and other relaxation mechanisms, such as external relaxation, are considered to be negligible.

Equation (1) can be solved as (Bodenhausen & Ernst, 1982):

$$\mathbf{A}(\tau_m) = \mathbf{\chi} \cdot \exp(-\mathbf{\lambda}\tau_m) \cdot \mathbf{\chi}^{-1} \cdot \mathbf{A}(0) \qquad (5)$$

where $\mathbf{\chi}$ is the matrix of eigenvectors of the relaxation matrix $\mathbf{R}$, $\mathbf{\lambda}$ is the diagonal matrix of eigenvalues, and τ_m is the mixing time. Calculation of the NOE involves evaluating the dipolar interactions between all 185 non-exchangeable protons of the L10 DNA decamer. The computational time required to diagonalize the complete relaxation rate matrix $\mathbf{R}$ to obtain the eigenvector and eigenvalue matrices of equation 5 is therefore quite long (about 65 seconds on an Amdahl 5680 mainframe). Since this time is approximately proportional to the dimension of the matrix cubed, a two-fold reduction in time is made by creating proton subsets, calculating cross-relaxation from protons within an 8 Å sphere to a proton of interest and including spin diffusion within the sphere. An additional factor of about two can be saved by using the minimal number of spheres needed to cover all observed crosspeaks. For example, if a sphere centered on i is used to calculate A_{ij}, no sphere is necessary on spin j since matrix $\mathbf{A}$ is real and symmetric ($A_{ij} = A_{ji}$). The problem is reduced from diagonalizing a 185 by 185 matrix to about 85 25 by 25 matrices (approximately 25 protons within a given 8 Å sphere), retaining the accuracy of the NOE calculation. Test calculations satisfactorily reproduce NOE build-up curves from Borgias & James (1988).

Alternatively, the NOESY intensities can be calculated from (Macura & Ernst, 1980):

$$\mathbf{A}(\tau_m) = \exp(-\mathbf{R}\tau_m) \cdot \mathbf{A}(0) \qquad (6)$$

and the exponential term can be expanded in a Taylor series:

$$\exp(-R\tau_m) = 1 - R\tau_m + 0.5\,R^2\tau_m^2 + \ldots \quad (7)$$

and NOE peak intensities may be calculated as:

$$A_{ij} = \delta_{ij} - R_{ij}\tau_m + 0.5\sum_k R_{ik}R_{kj}\tau_m^2 + \ldots \quad (8)$$

This series does not converge rapidly for large molecules that tumble slowly in solution with greater cross-relaxation rates and when the mixing time is long. The matrix analysis procedure is not only generally applicable, but also is often computationally more efficient.

C. Distance determination procedures

In order to emphasize methods of distance determination without the problems of low signal to noise ratios and spectral artifacts in the NOESY experiment (Williamson & Neuhaus, 1987), three procedures are considered with calculated NOEs as idealized data. NOE intensities are simulated using the matrix analysis calculation for various proton pairs from a model of L10 in a B-DNA conformation (Arnott & Hukins, 1972) with a correlation time of 3 nanoseconds and a spectrometer frequency of 500 MHz. The distance determination methods are applied to the data calculated for mixing times of 50, 100, and 150 milliseconds. Reflecting the experimental conditions, the methods are evaluated for reliability and their ease of implementation.

1. Methods of distance determination using calculated NOE intensities

Method 1. At a sufficiently short mixing time, the NOESY cross-peak intensity is represented by the term linear in τ_m of equation 8, which is directly proportional to the inverse sixth power of the distance separating a pair of protons. However, the intensity of any given NOE crosspeak can be affected by spectral artifacts, especially in spectra with low signal to noise ratios and with possible contributions from zero quantum coherence at short mixing times. Therefore, NOE intensities are commonly measured as a function of mixing time with the slope of a line starting at zero and going through the NOEs determining a build-up rate, K. To determine distances, initial NOE build-up rates are typically calibrated with a reference build-up rate of a proton pair of known separation:

$$K_{ij} = K_{ij(ref)} \left(\frac{r_{ref}}{r_{ij}} \right)^6 \quad (9)$$

Figure II-1 shows buildup curves based upon calculated NOE data. In Figure II-1A, the initial rates are indicated by straight lines fitted through NOEs at 50, 100, and 150 milliseconds, except for the 2', 2" proton pair where the intensity past 100 milliseconds is not used since the non-linearity of the build-up curve indicates that the isolated spin pair approximation is inappropriate. This is not apparent for NOE curves of lower intensity, especially in the presence of some spectral noise. The cytosine base H5-H6 NOE with a distance of 2.46 Å is used for calibration of build-up rates in order to estimate the other interproton distances.

Method 2. The build-up of the NOE with τ_m can be fitted with a second order polynomial (equation 8). This procedure uses the term linear in τ_m of the 2 term fit to the NOE data to represent the initial build-up rate of the NOE (Hyberts & Wagner, 1989). Calibration of the initial rates using the reference leads to distances that have had some correction for spin diffusion. The two-term polynomial representation better reproduces the NOE data in Figure II-1A up to a mixing time of 150 milliseconds. Evaluation with yet a higher order polynomial is most often not feasible experimentally because of inaccuracies in determining cross-peak intensities and restrictions on the number of time points that can be collected during the available spectrometer time. Build-up rates from Methods 1 and 2 are summarized in Table II-1.

Table II-1 Initial NOE buildup rates (s^{-1}) of selected proton pairs for L10 DNA

	Proton pairs			
	A_4 2'<->2"	C_6 H5<->H6	A_4 1'<->T_5 H6	A_4 H8<->T_5 H6
method 1	158	48	20	2.6
method 2	388	48	31	0.9

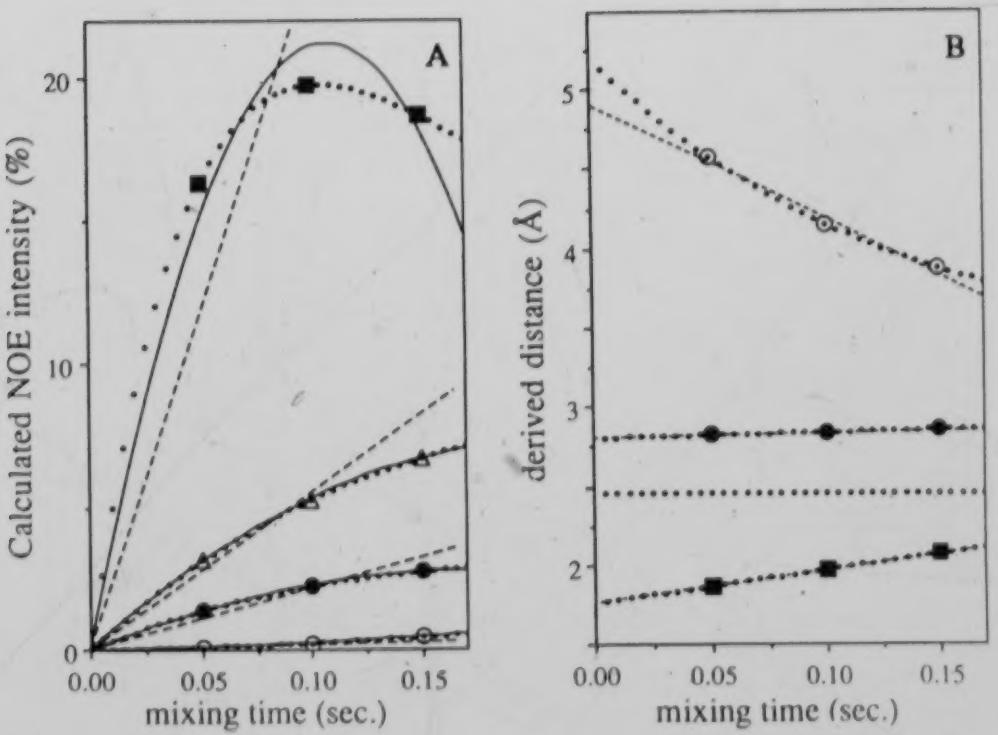


Figure II-1 Evaluation of distance determination methods based on simulated NOEs

NOE intensities are simulated at mixing times of 50, 100, and 150 milliseconds for: A_4 2'<->2" (■); C_6 H6<->H5 (Δ); A_4 1'<->T_5 H6 (●); and A_4 H8<->T_5 H6 (o) in a central portion of L10 in a B-DNA conformation. In A, the dotted lines represent the calculated NOE build-up curves. Initial build-up rates for a single term fit (isolated spin-pair approximation) are illustrated by dashed lines. The two-term fit is shown with a solid line. In B, distances are derived at each mixing time by comparing the NOE intensity to that of the reference. The horizontal line represents the reference cytosine H6-H5 at 2.46 Å. Linear extrapolation of derived distances to zero mixing time gives an estimate of the distance in each proton pair.

Method 3. A third procedure is the distance extrapolation method (Figure II-1B). The NOE for a given proton pair is compared to the reference NOE at each mixing time and the distance is derived at that mixing time:

$$r_{ij} = \lim_{\tau_m \to 0} r_{ij}(\tau_m) = \lim_{\tau_m \to 0} \left\{ \frac{NOE_{ref}(\tau_m) * r_{ref}(\tau_m)^6}{NOE_{ij}(\tau_m)} \right\}^{\frac{1}{6}} \qquad (10)$$

Distances at short mixing times are most accurate, and deviate uni-directionally from the correct value at long mixing times because of spin diffusion. Therefore, extrapolation to zero mixing time gives an estimate of the interproton separation. The method is similar to Figure 2 of Nilges et al. (1987), except that they define a reference NOE intensity as the product of an initial slope estimate and the mixing time. In Figure II-1B, the tendency to similar NOEs and therefore calculated distances is seen at long mixing times where spin diffusion is important. A positive slope indicates a loss of magnetization to other spins faster than the reference cytosine H5-H6 pair, whereas a negative slope indicates a relative gain in cross-peak intensity via spin diffusion. A polynomial fit through the distances would increase the accuracy of the calculation, although insufficient data points are taken to warrant such a procedure experimentally.

The results for interproton distances derived by these methods are summarized in Table II-2. The distance extrapolation method is clearly better than the 1-term method and is generally as good as distances derived from the NOE build-up rate approximated by the linear term of a two term fit to the build-up curve. Since the distance extrapolation method is also self-correcting for changes in instrument gain between NOESY experiments of different mixing time and more directly demonstrates the effects of spin diffusion on derived distances, it is a method of choice, at least for the proton pairs represented in Figure II-1.

Table II-2 Distance estimates (Å) based on simulated NOE data for L10 DNA

	Proton pairs			
	A_4 2'<->2"	C_6 H5<->H6	A_4 1'<->T_5 H6	A_4 H8<->T_5 H6
Correct distance	1.76	2.46	2.80	5.20
method 1	1.95	(2.46)	2.85	3.99
method 2	1.85	(2.46)	2.81	5.09
method 3	1.77	(2.46)	2.80	4.89

In Figure II-2, distances estimated from the three procedures are compared to the correct distances found for all the intra-residue and inter-residue proton pairs that are commonly observed by NMR in the NOESY spectrum of a B-type DNA duplex[4]. Distances longer than the reference 2.46 Å are systematically underestimated, especially with the isolated spin-pair approximation method. The two poorest distance estimates are near 3.8 Å. These are the intranucleotide aromatic base<->2" and the internucleotide 2'<-> aromatic base (n+1) proton pairs. They represent the near linear arrangements of protons spaced about 2 Å apart for the aromatic base proton, the methylene 2' and 2" sugar protons of that nucleotide(n), and the base proton of the succeeding 3' (n+1) nucleotide. This geometry is most troublesome for analysis of NOE data (Landy & Nageswara Rao, 1989).

2. Distance and initial structure determination using observed NOE intensities

The distance extrapolation method is now used on real experimentally measured NOE data to derive interproton distances which then form a basis set for structural determination. NOESY spectra were obtained at 20°C on a Varian VXR-500 spectrometer for a 4 mM sample of L10 at average mixing times of 50, 100, 150, and 250 milliseconds with a randomization of ±10 milliseconds to suppress quantum coherences. Volume integration of cross-peaks up to 150 milliseconds gave the NOE intensities used for distance determination, except for the longest distances for which an NOE could not be observed at shorter mixing times. For these proton pairs, a distance estimate was made by also considering the derived distance at 250 milliseconds. In Figure II-3, NOE build-up curves and distance extrapolations are illustrated for the same proton pairs as were used above to illustrate the behavior of the distance determination methods. Because the experimental structure is different than canonical B DNA, the A_4 1'<->T_5 H6 distance, for example, appears longer. Upper and lower bounds on the distances are estimated from assuming an error of approximately 20% in the NOE measurement (see below).

292 experimentally determined distances are used in a restrained molecular dynamics (MD) simulation to generate a structure consistent with the distance restraint set. The calculations were performed with the GROMOS program (Kaptein et al., 1985; de Vlieg et al., 1986), following a procedure for oligonucleotides (Nilges et al., 1987), that will be

---

[4]Intranucleotide (for Adenine 4): 1'<->2", 2.30 Å; 1'<->2', 2.98 Å; 1'<->3', 3.89 Å; 1'<->4', 3.60 Å; 1'<->H2, 4.51; 1'<->H8, 3.86 Å; 2'<->2", 1.76 Å; 2'<->3', 2.37 Å; 2'<->H8, 2.16 Å; 2"<->3', 2.69 Å; 2"<->H8, 3.66 Å; 3'<->4', 2.70 Å; 3'<->5', 3.71; 3'<->5", 2.81; 3'<->H8, 4.15 Å; 4'<->H8, 4.80 Å; and internucleotide (i<->i+1; Adenine 4 - Thymine 5): 1'<->H6, 2.80 Å; 1'<->1', 4.94 Å; 2'<->H6, 3.87 Å; 2"<->H6, 2.36 Å; 3'<->H6, 4.97 Å; H8<->H6, 5.20 Å.

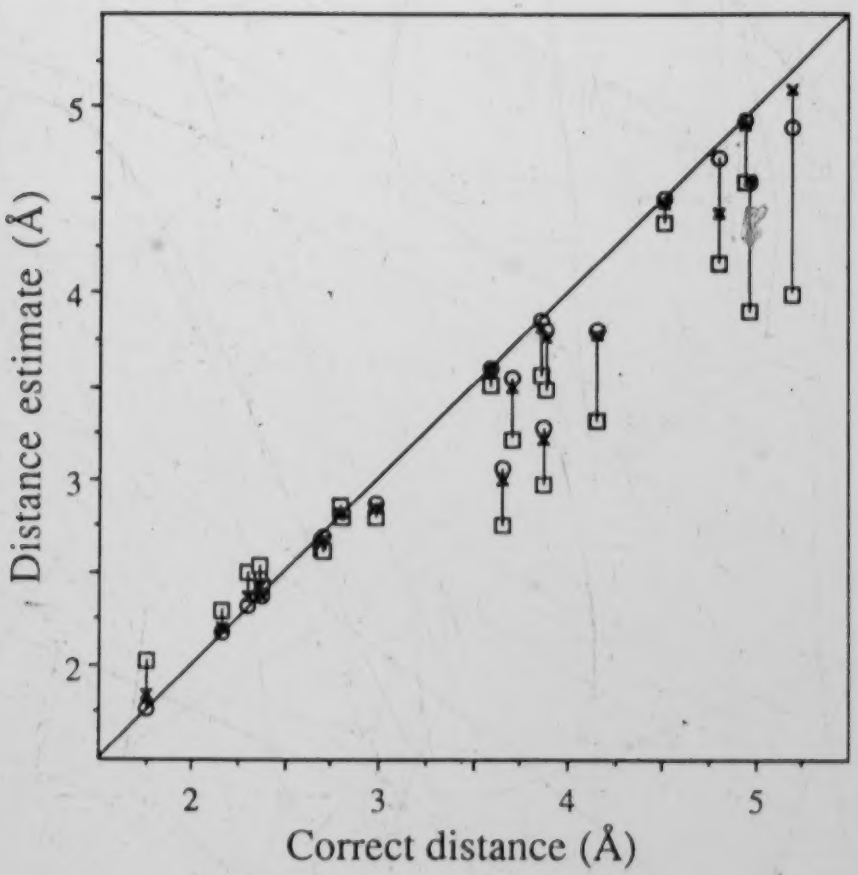


Figure II-2 Distance estimate comparisons for L10 in a B-DNA conformation
NOE intensities are first calculated using a full matrix analysis procedure. Distances estimated from the simulated NOE intensities using the three distance determination methods are compared to the correct distances for proton pairs commonly observed in the NOESY spectrum of a DNA duplex in a B-DNA conformation. Intra-residue proton pairs are illustrated for $Adenine_4$ and inter-residue pairs are represented by the $Adenine_4$ <-> $Thymine_5$ base step: Initial build-up rate from a single term fit to the NOE build-up curve (□); Term linear in τ_m from a two-term fit, $A\ \tau_m + B\ \tau_m^2$, (x); Derived distance extrapolation (o).

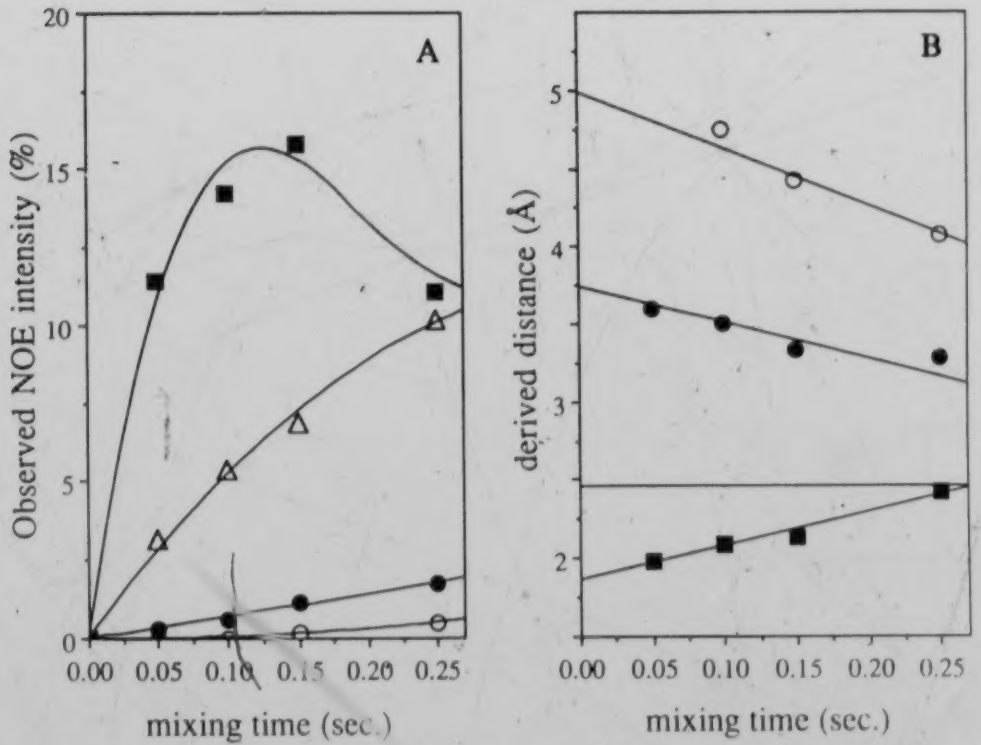


Figure II-3 Distance determination from experimentally observed NOEs in L10
The measured NOE intensities up to a mixing time of 250 milliseconds are presented in A for A_4 2'<->2" (■); C_6 H6<->H5 (Δ); A_4 1'<->T_5 H6 (●); and A_4 H8<->T_5 H6 (o). In B, distances are derived at each mixing time by comparing the NOE intensity to that of the reference (C H6-H5, 2.46 Å). Linear extrapolation of derived distances to zero mixing time gives an estimate of the distance in each proton pair.

more completely described in Chapters III and IV. For the purposes of this discussion, we consider results from a 20 picosecond MD run starting from canonical A DNA. In the first half of the MD run, force parameters were increased from 500 to 10000 $kJ \cdot mol^{-1}\ nm^{-2}$ for the distance restraints. Averaging coordinates over the last 5 picoseconds the MD run yields MD structure I.

Molecular dynamics procedures were chosen over other structure generating methods (e.g. distance geometry, Havel & Wüthrich, 1985) since the MD programs were readily available and most completely incorporated general concepts of molecular forces such as electrostatic and van der Waals forces. Although approximate, inclusion of a MD force field should give more accurate structures than the neglect of the interatomic interactions, especially when the number of observables from NMR is small (Chapter III, Nilges et al., 1987).

D. NOE-based refinement

In the second part of this manuscript, MD structure I is refined by comparing the experimentally observed NOEs to those calculated from the structure using the matrix analysis procedure. Since experimentally observed NOE cross-peaks are measured in arbitrary units, they are first scaled to calculated NOE intensities by multiplying with a scaling factor ($\Sigma NOE_{obs} / \Sigma NOE_{calc}$). NOE-based refinement is accomplished by replacing the distance target function in the total energy description of the decamer DNA system with an NOE potential term:

$$E_{NOE} = 0.5 * CNOE * \Sigma [\, NOE_{obs} - NOE_{calc} \,]^2 \qquad (11)$$

where CNOE is a force constant set equal to 2000 $kJ \cdot mol^{-1}\ (\Delta NOE)^{-2}$. Forces from the NOE potential are evaluated at the first step and are subsequently evaluated every ten steps during the molecular mechanics run. Spheres are again used, first to select all protons within 8 Å of a given proton i. The force on proton i is then calculated using only observed NOEs involving protons within the 8 Å sphere. Derivatives of the NOE with respect to a change in each Cartesian coordinate of i are evaluated numerically by changing the coordinate slightly (by 0.01 Å) and re-calculating the NOE:

$$F_{x_i} = -\frac{\partial E_{NOE}}{\partial x_i} = CNOE * \Sigma [\, NOE_{obs} - NOE_{calc} \,] * \frac{\partial NOE_{calc}}{\partial x_i} \qquad (12)$$

Evaluation of the function is repeated for each proton in the molecule. These NOE 'energy' gradients include both direct forces, where two protons, i and j, are pushed closer together if the calculated NOE_{ij} is too weak, or stretched apart if too strong, and indirect forces that result from the effect of moving proton i on NOE_{jk}. Note that although proton i may not be observed at all as an NOE, information on its position is known because of the effect on nearby NOEs (Landy & Nageswara Rao, 1989; Fejzo et al., 1989). Here, the pseudo-energy forces are calculated for NOE intensities between non-exchangeable protons for mixing times of 150 and 250 milliseconds, although, if desired, the algorithm (Appendix 1) could be easily extended to take into account all mixing times.

A residual factor R monitors the fit of the NOEs calculated from the structure to observed NOE intensities:

$$R = \frac{\Sigma | NOE_{obs} - NOE_{calc} |}{\Sigma NOE_{obs}} \qquad (13)$$

where the summation runs over the number of observables. To keep an analogy with the standard crystallographic R factor, weighting factors based on the standard deviation in the NOE intensity are not included. R factors are presented in Table II-3 for mixing times of 150 and 250 milliseconds which is a compromise between less signal-to-noise at shorter mixing times, and greater spin diffusion effects at long times. The observed NOE intensities are tabulated for all assigned, non-overlapping NOE cross-peaks and are here divided into three groups: (1), strong NOEs that correspond to distances of less than about 2.8 Å; (2), strong and medium NOEs for distances of less than about 3.6 Å; and (3), all NOE intensities. Table II-3 indicates higher R values for data that includes the weaker intensities. After 100 steps of energy minimization, the R factor is reduced to 0.19 (for all observed NOEs at mixing times of 150 and 250 milliseconds), indicating a close agreement between structure and observed data. The refined structure II has an overall atomic rms deviation of 1.3 Å from canonical B DNA and is depicted in Figure II-4.

Table II-3 R factors for observed NOE intensities of L10

	strong NOEs	strong and medium NOES	all NOEs
# of observations[a]	234	394	603
A DNA model	0.43	0.61	0.65
B DNA model	0.31	0.38	0.42
MD structure I[b]	0.22	0 26	0.27
MD structure I[c]	0.19	0.23	0.24
Refined structure II[d]	0.15	0.18	0.19

[a] 294 NOE intensities in total were observed at a mixing time of 150 milliseconds and 309 intensities at 250 milliseconds. [b] Structure produced after 20 picoseconds of molecular dynamics restrained with approximate distances derived by extrapolation to zero mixing time. [c] MD structure I with correlation time reduction factors. [d] Structure after 100 cycles of energy minimization with observed NOEs and including differential motion.

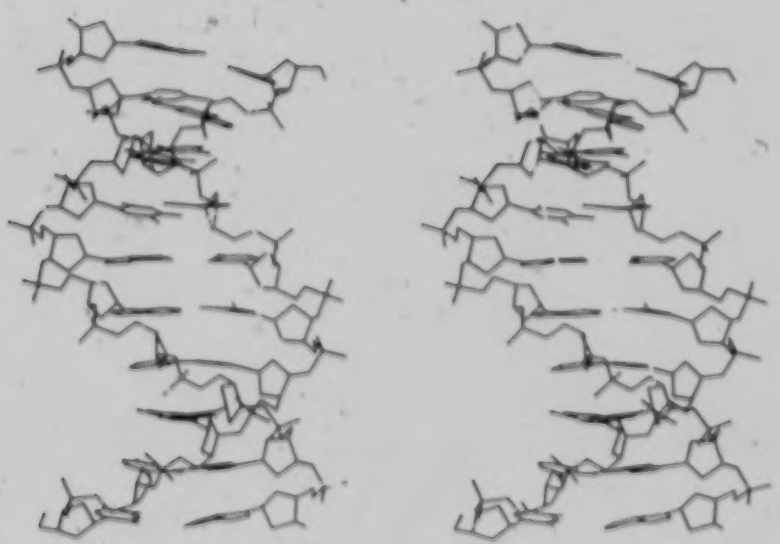


Figure II-4 Stereoview of the refined structure of L10 DNA

Calculation of the limiting R factor to be expected for a known structure was determined by first assessing the experimental uncertainties in observed NOE intensities. Two sets of NOE intensities were available from two 50 millisecond mixing time NOESY spectra of L10 which had been collected several days apart. The absolute value of the difference for each cross-peak intensity in the two sets increase with the size of the cross-peak (Figure II-5), and this relative error can be estimated as about 20 % of the cross-peak intensity. The relative error represents effects, such as t_1 and t_2 streaking and baseline distortion, that are approximately proportional to the cross-peak intensity. In addition, there is an absolute error component of 15 (in arbitrary units of measured NOE intensities) which represents the random noise for the NOESY spectrum.

NOE intensities were calculated for L10 at 150 and 250 milliseconds mixing time in a known structure, namely that of a B-DNA conformation. Random errors were added[5] to these NOEs with the standard deviations set on a Gaussian distribution function assuming a relative error of 20%, and an absolute error contribution of 15 (i.e., $0.20*NOE_{calc} + 15$). The R factor for the B-DNA NOE intensities with added errors is 0.17, which represents a limit at which the structure is consistent with the NOE data.

During refinement of the structure, correlation time adjustment factors were assigned to each proton in the DNA molecule. These are used to locally reduce correlation times for certain proton pairs and introduce a differential motion model for DNA:

$$\tau_{ij} = \tau_c * (S_i * S_j) \qquad (14)$$

The product, $(S_i*S_j) = S_{ij}^2$, relates to the order parameter, S^2, of Lipari and Szabo (1982), and can vary between 0 and 1. A value of one indicates that the correlation time of the interproton vector is the same as the overall tumbling time of the macromolecule; a value of zero indicates complete motional freedom. Most protons have an adjustment factor of one. Reduced correlation times might be expected for groups of protons associated with local internal motion in DNA, such as thymine methyl rotation, and flexibility about the sugar 2' methylene. Additional motion should be expected for protons of the 5' and 3' terminal residues (Clore & Gronenborn, 1984; Hogan & Jardetzky, 1980; Keepers & James, 1982).

---

[5] NOE values with random errors of Gaussian distribution were generated with the desired standard deviation using the GAUSS subroutine of the GROMOS molecular dynamics simulation program package.

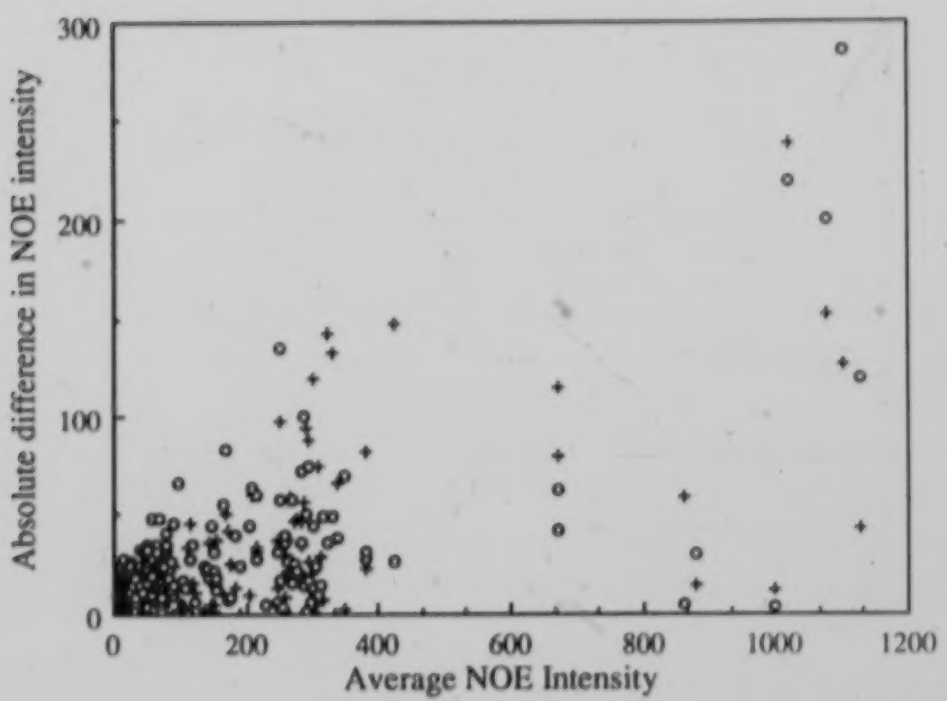


Figure II-5 Error estimation in observed NOE intensities

Two separate 50 millisecond mixing time NOESY experiments were taken several days apart. Cross-peak intensities for both were measured and tabulated. The NOE intensities were corrected for changes in instrument gain and sample over the measurement period. The first set of intensities was overall approximately 8% lower than the second set. The absolute difference between the first and second measurement (after correction) is plotted against the average NOE intensity for each measured cross-peak (o). Errors were estimated to have a relative error of approximately 20% and an absolute error of 15 units (when scaled to calculated NOE intensities, an absolute error of 0.005). Assuming these values, a corresponding set of random errors was constructed on a Gaussian distribution for the averaged NOE intensity. The absolute value of these errors is plotted against the averaged NOE intensity for each cross-peak (+).

The addition of correlation time adjustment factors do not take into account that the NOE between protons not fixed by covalent geometry can be modified by distance fluctuations. The observed cross-peak is an approximate r^6 weighted average, and is therefore dominated by the close approach of the atoms. Therefore, the incorporation of differential correlation times is first examined for proton pairs of fixed length (Figure II-6). These include the covalently bound proton pairs (the methylene 2'<->2", 1.76 Å and the cytosine H5<->H6, 2.46 Å) and proton pairs of nearly invariant length (the sugar ring 1'<->2", which is 2.3±0.1 Å for all sugar ring conformations and the thymine H6<->CH_3 (center), 3.0 Å). To improve signal-to-noise in Figure II-6, crosspeak intensities (excepting the 5' and 3' termini) were averaged for each type of proton pair (2'<->2", H5<->H6, 1'<->2", and H6 <->CH_3).

In Figure II-6A, NOE intensities are calculated using the same correlation time, 2.8 nanoseconds, for all interproton vectors. The calculated values for the average 2'<->2" and thymine H6<->CH_3 NOE intensities are clearly too large and the cytosine H5<->H6 and sugar 1'<->2" intensities too low. These errors are larger than those expected for small errors in the internuclear distances used. The NOE R factor (over individual NOE intensities) is 0.19. The NOE for the corresponding H6<->CH_3 cross-peak is merely the sum of the three individual H6<->methyl proton crosspeaks, with the methyl group in a fixed conformation. Since methyl groups often nearly freely rotate, the neglect of internal motion here is clearly inappropriate (Rowan et al., 1974). Internal motion has also been neglected in Figure II-6A for motion of the 2' and 2" protons of DNA (Lefèvre et al., 1987; Clore & Gronenborn, 1984).

In Figure II-6B, NOE intensities are re-calculated using correlation time adjustment factors allowing different correlation times for each class of proton pair. The best fit is given with a slower correlation time for the cytosine H5<->H6 vectors and faster correlation times for the remainder of the interproton vectors. Here, NOE intensities are calculated using correlation time adjustment factors of 0.5 and 0.8 for the thymine methyl and methylene protons on the 2' carbon, respectively, and 1.0 for the other protons, with a molecular correlation time, τ_c, of 3.8 nanoseconds. The average interproton vector correlation time is same as before—2.8 nanoseconds. The NOE R factor is reduced to 0.11 by allowing differential motion.

The calculated NOE intensities still do not match observed NOE intensities well in Figure II-6B for the average thymine H6<->CH_3 crosspeak. Although inclusion of a shorter correlation time for the interproton vector is convenient, a more sophisticated model, such as jumping between allowed methyl rotor states (Rowan et al., 1974), could be more accurate. The agreement between observed and calculated 1'<->2" intensities

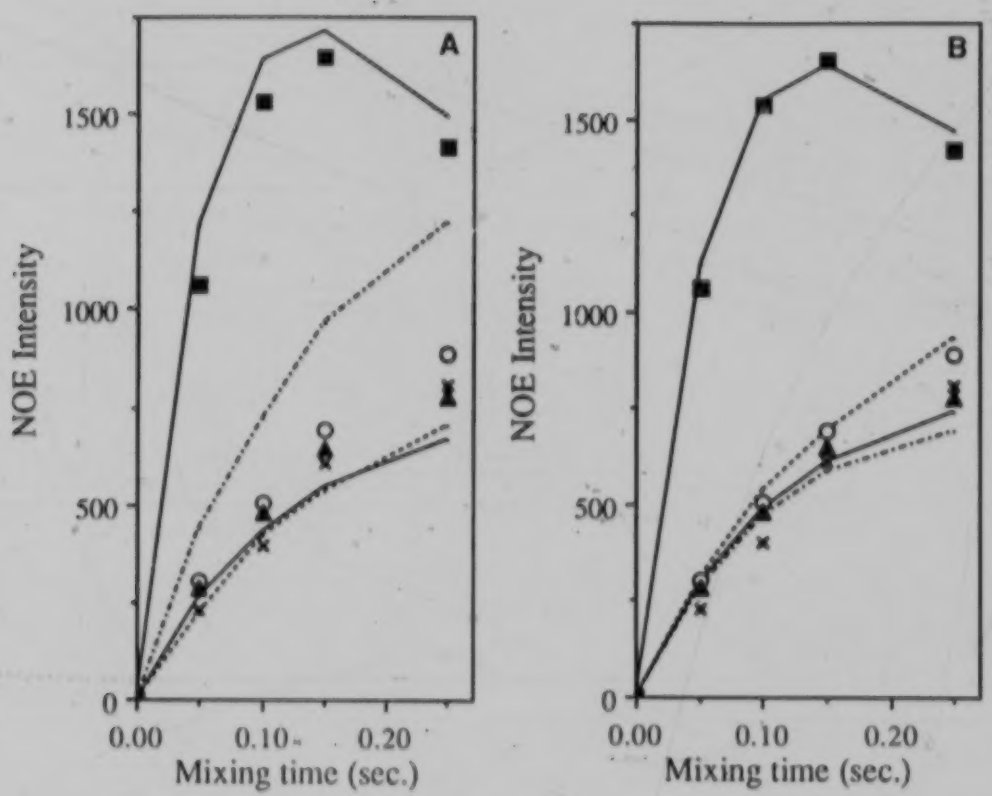


Figure II-6 Nuclear Overhauser effect build-up curves for proton pairs of invariant length Volume integrals of two-dimensional NOE cross-peaks were averaged for each class of proton pair: sugar ring 2'<->2" (■), cytosine H5<->H6 (o), sugar ring 1'<->2' (▲), thymine H6<->CH_3 (x). The lines show NOE intensities calculated using a full matrix analysis procedure: sugar ring 2'<->2" (—, upper), cytosine H5<->H6 (- - -), sugar ring 1'<->2" (—, lower), thymine H6<->CH_3 (····). For ease of comparison between the panels, calculated NOE intensities are fit to the observed NOE intensities although, normally, observed NOE intensities are instead scaled to those calculated. In Panel A, NOE cross-peaks are computed using the same correlation time, 2.8 nanoseconds, for all interproton vectors. A differential motion model is used to calculate NOE intensities in panel B resulting in the following correlation times (in nsec.): sugar ring 2'<->2" (2.4), cytosine H5<->H6 (3.8) , sugar ring 1'<->2" (3.0), thymine H6<->CH_3 (1.8).

could be improved by choosing different adjustment factor components for the 2' and 2" protons (0.85 and 0.75, respectively, instead of 0.80 for both), although the improvement would be small, and the distinction in motion between the 2' and 2" protons difficult to justify on physiochemical grounds. Other small deviations between observed and calculated intensities most likely result from experimental errors in the observed intensities, such as those caused by limited digital resolution and spectral overlap, and from errors arising by the neglect of incomplete relaxation between scans and of non-dipolar sources of relaxation such as external relaxation caused by paramagnetic ions. In addition, fluctuations in atomic positions for protons that have an indirect influence on the covalent proton pairs (such as movement of the 3' protons) are not included. Anisotropic motion likely plays no significant role in modifying NOE intensities since a DNA duplex ten base-pairs long is nearly gobular (approximately a cylinder 30 Å long and 24 Å in diameter) and the improvement in the NOE R factor upon incorporation of the differential motion model is noted to be more or less equal for more globular DNA octamer duplexes (Chapter III).

The inclusion of correlation time adjustment factors also improves the overall comparison between observed and calculated NOE intensities for the 150 and 250 millisecond mixing time NOE intensities used for structure refinement. Here, values of 0.65, 0.85 and 0.9 are used empirically to reflect the increased motion of all thymine methyl groups, sugar 2' and 2" methylene protons, and the 5' and 3' terminal residues, respectively (Clore & Gronenborn, 1984; Keepers & James, 1984), with an overall correlation time of 3.8 nanoseconds[6]. The incorporation of order parameters for MD structure I improves the NOE R factor from 0.27 to 0.24 (Table III-3).

The local correlation time reduction factors improve the comparison between observed and calculated NOE intensities mainly by taking into account fast motions that locally reduce the rotational correlation time for an interproton vector. Methods treating the observation of slower motional processes by NMR have been recently reviewed (van de Ven & Hilbers, 1988; Williams, 1989). Longitudinal T_1 and transverse T_2 relaxation times have been traditionally measured to demonstrate internal mobility in macromolecules. The homonuclear NOE may also be used not only to determine the structure of a molecule in solution, but also to quantitatively describe some of its dynamics of motion.

---

[6]In the absence of correlation time adjustment factors, the best agreement between observed and calculation NOE intensities for the 50-250 millisecond NOEs between protons in fixed geometries occurs with a molecular correlation time, τ_c, of 2.8 nanoseconds. For all 150 and 250 millisecond NOE intensities used for structure refinement the best fit is with a τ_c of 3 nsec. Likewise, the correlation time adjustment factors are slightly different for the two groups of NOE intensities.

A more sophisticated model of motion could be used (Noggle & Shirmer, 1971; Lipari & Szabo, 1982; Keepers & James, 1984; Rowan et al., 1974), but the incorporation of model-free empirical parameters here is sufficient with respect to the precision of the NOE data experimentally collected. The incorporation of empirical correlation time reduction factors can be viewed as analogous to the addition of thermal factors used in refinement of structures from X-ray crystallographic data.

Although it may be preferable to use NOE intensities directly in a long restrained MD run, thereby performing a better search of conformational space and eliminating the step of making rather approximate distance determinations, the computational time would be cumbersome. Much more time would be required on the Amdahl (approximately 30 hours) when compared to the two hours needed for a molecular dynamics run restrained with distances. Moreover, it is unclear whether the structure produced by using the NOE directly in MD runs would be essentially more correct than that produced using approximate distances, followed by refinement with NOEs, given any inadequacies in the motional model used here.

Full use of NOE data will be facilitated by further improvements in calculational strategy (Yip & Case, 1989), a more complete description of molecular motion, and by the availability of faster computers. Often at long mixing times, NOE cross-peaks can be observed between protons even 6 and 7 Å apart, but this information is lost since it cannot be directly related to a distance. These NOE calculation methods indicate a way to obtain more parameters that cover a greater range of conformation space, and to better determine structures in solution by NMR.

E. References

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Oxford University Press, London.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504-1509.

Baleja, J. D., & Sykes, B. D. (1988) *Biophys. J. 53*, 80a.

Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 1304-1309.

Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290-308.

Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493-512.

Bull, T. E. (1987) *J. Magn. Reson. 72*, 397-413.

Clore, G. M., & Gronenborn, A. M. (1984) *FEBS Letters 172*, 219-225.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Israel J. Chem. 27*, 181-188.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two dimensions*, Oxford University Press, Oxford.

Fejzo, J., Zolnai, Z., Macura, S., & Markley, J. L. (1989) *J. Magn. Reson. 82*, 518-528.

Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *Biochemistry 27*, 7909-7919.

Harris, R. K. (1983) *Nuclear Magnetic Resonance Spectroscopy*, Pitman, London.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281-294.

Hogan, M. E., & Jardetzky, O. (1980) *Biochemistry 19*, 3460-3468.

Hyberts, S. G., & Wagner, G. (1989) *J. Magn. Reson. 81*, 418-422.

Kaptein,R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179-182.

Keepers, J. W., & James, T. L. (1982) *J. Am. Chem. Soc. 104*, 929-939.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404-426.

Landy, S. B., & Nageswara Rao, B. D. (1989) *J. Magn. Reson. 83*, 29-43.

Lefèvre, J.-F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076-5090.

Lipari, G., & Szabo, A.(1982) *J. Am. Chem. Soc. 104*, 4546-4559.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95-117.

Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987) *Biochemistry 26*, 3734-3744.

Noggle, J. H., & Shirmer, R. E. (1971) *The Nuclear Overhauser Effect*, Academic Press, New York.

Olejniczak, E. T., Gampe, Jr., R. T., & Fesik, S. W. (1986) *J. Magn. Reson. 67*, 28-41.

Rowan, R., McCammon, J. A., & Sykes, B. D. (1974) *J. Am. Chem. Soc.96*, 4473-4780.

Slichter, C. P. (1978) *Principles of Nuclear Magnetism*, 2nd edition, Springer-Verlag.

Solomon, I. (1955) *Phys. Rev. 99*, 559-565.

Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854-6865.

van de Ven, F. J. M., & Hilbers, C. M. (1988) *Eur. J. Biochem. 179*, 1-38.

Williams, R. J. P. (1989)*Eur. J. Biochem. 183*, 479-497.

Williamson, M. P., & Neuhaus, D. (1987) *J. Magn. Reson. 72*, 369-375.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Yip, P., & Case, D. A. (1989) *J. Magn. Reson. 83*, 643-648.

Chapter III

The Solution Conformation of Purine-pyrimidine DNA Octamers using Nuclear Magnetic Resonance, Restrained Molecular Dynamics, and NOE-based Refinement

A. Introduction

Although DNA displays considerable structural diversity, deoxyoligonucleotides can be classified into one of several distinct conformational families. The relative stability of possible conformations is dependent upon both base sequence and external factors such as solution media composition and the presence of superhelical torsion forces. NMR experiments (Cohen, 1987; Gronenborn & Clore, 1985; Patel et al., 1982), crystallographic determinations (Wang et al., 1979, Dickerson & Drew, 1981), and enzymatic studies (McLean & Wells, 1988; Naylor et al., 1986) of the right-handed A and B-type helical forms, and the left-handed Z form exemplify the heterogeneity within each major structural class. Z conformations are found in sequences with a regular alternation of purine and pyrimidine residues which are mostly guanine and cytosine, although there have been some exceptions (Feigon et al., 1985; McLean & Wells, 1988). Alternating adenine-thymine tracts, however, adopt right-handed conformations in solution (Lomonossoff et al., 1981; Suzuki et al., 1986). Intermediary sequences of 50% GC content are of interest not only because these sequences appear in genomes, such as in the anti-Z DNA antibody binding regions of SV40 viral DNA (Hagen et al., 1985; Nordheim & Rich, 1983), but also because of the exhibition of conformational polymorphism (Z-DNA tract formation versus cruciform extrusion) as a response to different conditions of superhelical stress when cloned into supercoiled plasmids (McLean & Wells, 1988; Naylor et al., 1988). To understand the predisposition of these sequences in promoting the B-Z transition, a determination of the structural details of $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ as linear duplexes without effects of superhelical stress was undertaken herein.

Two dimensional ^1H NMR techniques yield data that can provide high-resolution molecular structures in solution. In particular, the nuclear Overhauser effect results from the spatial proximity of nuclear spins and can be used to determine their separation (Noggle & Shirmer, 1971; Havel & Wüthrich, 1985; Patel et al., 1987). This data can be supplemented with measurements of coupling constants (J), which provide quantitative estimates of torsion angles (Rinkel & Altona, 1987; Chary et al., 1988). Sets of interproton and dihedral measurements are then used as a basis for structure determination by incorporation of these restraints as effective potentials into the total energy function of the system in restrained molecular dynamics simulations (Behling et al., 1987; Kaptein et al., 1985). The empirical energy function ensures that structures which satisfy the experimental restraints still have approximately correct local stereochemistry and non-bonded interactions (Nilsson al., 1986; Nilsson & Karplus, 1986).

The distance between two spins is often estimated from NOE data by assuming inverse proportionality to the sixth root of the NOE cross-peak intensity. Distances so derived are only approximate since the cross-peak intensity due to direct cross-relaxation between spins i and j is modified by additional cross-relaxation with any spin k, especially if spin k exists such that $r_{ik} < r_{ij}$ or $r_{jk} < r_{ij}$. However, NOE cross-peak intensities may be predicted from the structures produced by a molecular dynamics simulation and compared directly to the observed intensities, eliminating approximate distance calculation. The structures resulting from restrained molecular dynamics calculations can be refined in an iterative manner so as to minimize the difference between the two sets of NOEs (Chapter II, Baleja et al., 1990a).

In this chapter, the solution structures of two self-complementary alternating purine-pyrimidine DNA oligonucleotides $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ are investigated. After sequential assignment of all the nonexchangeable protons, approximate interproton distances are obtained by extrapolating distances derived from time-dependent NOE measurements to zero mixing time (Chapter II). Estimates of coupling constants from one and two-dimensional spectra enable us to limit the conformational space for all glycosidic torsion angles. For each oligonucleotide, restrained molecular dynamics simulations are started from A and from B DNA (atomic rms difference of 4.3 Å) and converge to structures that satisfy the experimental restraints. For $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$, the atomic rms difference between the averaged dynamics structures (0.66 and 0.65 Å, respectively) is comparable to the rms fluctuations of the atoms about their average positions. The resulting structures are refined by comparing observed NOE intensities to the NOE intensities calculated using a full matrix analysis procedure and minimizing the difference between the two sets of NOEs (Chapter II). Final structures have NOE R factors of 0.19 for $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and 0.23 for $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$, each of which represent a high quality of fit to the experimental data. Conformational parameters of the structures are analyzed with respect to base-sequence dependence and chain termination effects. In general, they reflect B-DNA type features, qualitatively in agreement with previous ^1H NMR studies of related sequences (Lown et al., 1984; Nilges et al., 1987; Nilsson et al., 1986; Searle et al., 1988; Stevens et al., 1988).

B. Experimental Procedures

1. Sample preparation

Deoxyoligonucleotides, 5' d(G-T-A-C-G-T-A-C) and 5' d(C-A-T-G-C-A-T-G), were prepared (by M. W. Germann & J. H. van de Sande, University of Calgary) on an Applied BioSystems Model 380A DNA synthesizer using phosphoramidate chemistry (Beaucage & Caruthers, 1981). After deblocking and detritylation, synthesis products were purified by anion-exchange chromatography at pH 13.0 on NACS-20 (Germann et al., 1985). Aliquots of the purified oligonucleotides, which were 5'-end labelled with [γ-^{32}P] adenosine triphosphate and T_4 polynucleotide kinase (Chaconas & van de Sande, 1980), gave a single band on 20% denaturing polyacrylamide gels for each product.

The purified oligonucleotides were de-salted (at the University of Alberta) by G25 chromatography and were lyophilized. By heating each sample in 1.3 ml of 10 mM Na·phosphate buffer, pH 7.0, 0.1 M NaCl, and 25 µM EDTA to 85°C, and allowing the solutions to cool to room temperature over several hours, strands annealed to form the duplexes:

5' (G_1 pT_2 pA_3 pC_4 pG_5 pT_6 pA_7 pC_8) 3' GTAC
3' (C_8 pA_7 pT_6 pG_5 pC_4 pA_3 pT_2 pG_1) 5'

and

5' (C_1 pA_2 pT_3 pG_4 pC_5 pA_6 pT_7 pG_8) 3' CATG
3' (G_8 pT_7 pA_6 pC_5 pG_4 pT_3 pA_2 pC_1) 5'

GTAC and CATG are abbreviations for [d(G-T-A-C-G-T-A-C)]$_2$ and [d(C-A-T-G-C-A-T-G)]$_2$, respectively. Solutions were passed over the Na^+ form of Chelex 100 to remove paramagnetic metal ions prior to lyophilization. Samples were dissolved and re-lyophilized three times in increasing grades of D_2O, and finally taken up in 0.65 ml of 99.997% D_2O. Product homogeneity was also checked by ^{1}H NMR spectroscopy (Figure III-1). Final buffer concentrations for both samples were 0.2 M NaCl, 20 mM Na·phosphate, pH 7.3 (direct meter reading), and 50 µM EDTA. Duplex concentrations were 0.5 and 0.6 mM for CATG and GTAC, respectively.

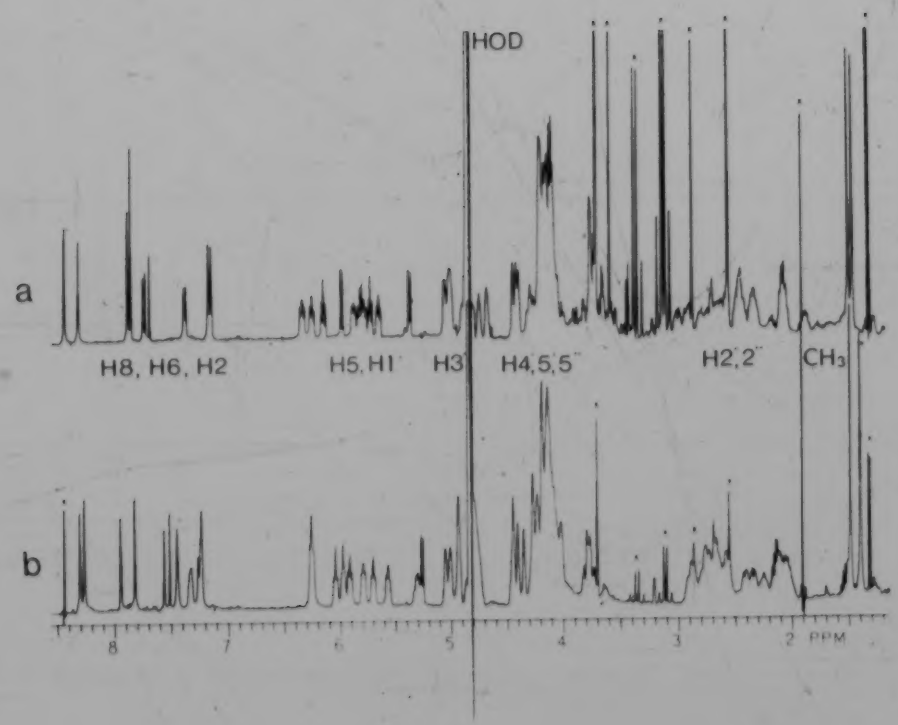


Figure III-1. One-dimensional NMR spectra of CATG and GTAC

One-dimensional 400 MHz ^1H NMR spectra are shown of: (a), $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$; and (b), $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$, in D_2O at 20°C. The spectral impurities, marked x, arise from small molecular weight impurities associated with oligonucleotide synthesis and purification, and did not interfere with two-dimensional spectral analysis.

2. Nuclear magnetic resonance spectroscopy

All NMR spectra were obtained on a Varian XL 400 NMR spectrometer with an operational frequency of 400 MHz for protons. Experiments were taken with 2K data points along t_2 and with 256 t_1 increments. The spectral width employed was approximately 3400 Hz, with a relaxation delay time of 2.0 seconds. Streaking along t_1 was reduced by multiplying the first domain time point by a factor optimized near 0.5 (Otting et al., 1986). Absolute value COSY spectra (Nagayama et al., 1980) were recorded at 26°. The appropriate phase cycling was used for quadrature detection and to eliminate axial peaks, and in the case of NOESY spectra, single and multiple quantum coherences.

Phase-sensitive NOESY spectra (States et al., 1982) were collected at 20°C, which was optimal for spectral resolution, and was a compromise between broader lines at lower temperatures and duplex fraying at higher temperatures. Average mixing times of 100, 200, 400, and 500 milliseconds were used. A random delay of between -10 and +10 milliseconds was incorporated to suppress zero quantum coherences. Spectra with the longer mixing times were mainly used for resonance assignment. Although spin diffusion was apparent, it did not interfere with the sequential assignment procedure. Low sample concentrations precluded the use of shorter mixing times.

Prior to two-dimensional Fourier transformation, the data were zero-filled to 2048 points along the t_1 dimension. COSY and NOESY data were weighted by both $\exp(t / RE)$ and $\exp(-t^2 / AF^2)$ functions in each dimension. Values of RE (to effect resolution enhancement) and AF (an apodization function to suppress truncation artifacts) were chosen so that for COSY spectra, the data were nearly nulled at the first and last time points (t), and were maximized at a point that corresponded to an acquisition time of about 1 / (2*J), where J is an average coupling constant (approximately 8 Hz.) best suited to observe most correlations. For phase-sensitive NOESY spectra, RE values were chosen to give slight resolution enhancement, and an AF value small enough to avoid truncation effects. Such a procedure reduces tailing about the diagonal with little change in cross-peak intensity. Final symmetrized two-dimensional spectra were 1K by 1K data points, representing a resolution of approximately 3.4 Hz per point. NOESY intensities were quantified by determining the volume integral of each cross-peak. Nominally empty areas perpendicularly adjacent to each cross-peak were examined for baseline correction.

3. Restrained molecular dynamics calculations

Energy minimization and molecular dynamics (MD) calculations were carried out with the GROMOS program (de Vlieg et al., 1986) and force field, which consisted of the usual terms for bonds, bond-angles, sinusoidal dihedral torsion, non-bonded interactions (van der Waals and electrostatics), and harmonic terms to maintain proper planar or tetrahedral geometries, and to which two extra terms representing distance and dihedral restraints were added. The distance restraint square well potential, E_{DIS}, was given by:

$$\begin{aligned} E_{DIS} &= 0.5 * CDIS * [\, r_{ij}^{U} - r_{ij} \,]^2 \qquad \text{if } r_{ij} > r_{ij}^{U} \\ &= 0.5 * CDIS * [\, r_{ij}^{L} - r_{ij} \,]^2 \qquad \text{if } r_{ij} < r_{ij}^{L} \end{aligned} \qquad (1)$$

where r_{ij}^{U} and r_{ij}^{L} were the upper and lower estimates of the distance between protons i and j, respectively, r_{ij} was the calculated distance, and CDIS was a force constant. The effective dihedral angle restraint, E_{DIH}, was represented by:

$$\begin{aligned} E_{DIH} &= 0.5 * CDLR * [\, \phi_k^{U} - \phi_k \,]^2 \qquad \text{if } \phi_k > \phi_k^{U} \\ &= 0.5 * CDLR * [\, \phi_k^{L} - \phi_k \,]^2 \qquad \text{if } \phi_k < \phi_k^{L} \end{aligned} \qquad (2)$$

where ϕ_k^{U} and ϕ_k^{L} were upper and lower allowed limits of the torsion angle, ϕ_k was the calculated angle, and CDLR was the force constant.

Several small alterations to the GROMOS force field were made in order to be more consistent with the nucleic acid force field of the CHARMM molecular mechanics programs (Nilsson & Karplus, 1986) and have been tested on a DNA decamer (Chapter IV; Baleja et al., 1990b). The normal van der Waals radius on united methylene carbons was reduced from 2.22 to 2.10 Å to avoid steric clashes between C2' and one of the oxygens on the 3' phosphate. Methine carbons were given van der Waals radii of 2.05 Å. Corresponding 1-4 van der Waals interactions were left unaltered. The effect of solvent was approximated for structure determination by a $1/\varepsilon r$ screening function where r was the separation of the charged groups in Å (Brooks et al., 1983) and the dielectric constant, ε was equal to 4 (Weiner et al., 1984). The net charge on each phosphate group was reduced to -0.32e (Tidor et al., 1983; Nilsson et al., 1986).

Starting models were first subjected to 200 steps of steepest descents energy minimization. During the first 10 picoseconds of each MD simulation, values of the distance restraint force constant were increased from 500 to 10000 $kJ \cdot mol^{-1}\ nm^{-2}$ and the

dihedral restraint force constant from 5 to 50 $kJ \cdot mol^{-1}$ rad^{-2}. Velocities were re-initialized (taken from a Maxwellian distribution at 300K) with every increase in the force constants (approximately every 1.5 picoseconds). Newton's equations of motion were integrated with a time step of 2 femtoseconds, with all bond lengths kept rigid by the SHAKE algorithm (Ryckaert et al., 1977). The molecule was weakly coupled to a temperature bath with a reference temperature of 300K using a coupling constant of 0.1 picoseconds (Berendsen et al., 1984). A cut-off radius of 10 Å was applied for non-bonded interactions and the nonbonded pair list was updated every 20 femtoseconds. Molecular dynamics runs with the highest values of CDIS and CDLR were continued to 20 picoseconds in total, and coordinates were averaged over the last 5 picoseconds. Averaged molecular dynamics structures were then subjected to 200 steps of energy minimization to correct distortions in the structure caused by the averaging procedure.

A total of four molecular dynamics runs for structural determinations with experimental distance and dihedral restraints were performed. The experimental data sets for each DNA octamer were based on observations from their two-dimensional NMR spectra (see below). For CATG, a MD run was started with a model in a classical A-type DNA conformation (Arnott & Hukins, 1972) and a second MD run used a starting model of CATG in an average B-DNA configuration. Two other MD runs were performed for GTAC, also with A and B-DNA starting models.

4. NOE-based structure refinement

Molecular dynamics calculations result in structures that satisfy the two sets of experimental restraints--the approximate distance set and the dihedral angle set. Distance restraints based on the NOE intensities are most often inaccurate because of spin diffusion effects. In addition, the distance restraints for Watson-Crick base-pairing forces some degree of planarity for the two bases of a base-pair, and may affect certain conformational parameters, such as propeller twist. Structures were therefore subjected to further refinement, directly based on the NOE intensities, by replacing the effective potential for distance restraints by the E_{NOE} potential for NOE restraints (Baleja et al., 1988; Chapter II):

$$E_{NOE} = 0.5 * CNOE * \sum [NOE_{obs} - NOE_{calc}]^2 \qquad (3)$$

NOE intensities between non-exchangeable protons were calculated from:

$$A(\tau_m) = \chi \cdot \exp(-\lambda \tau_m) \cdot \chi^{-1} \cdot A(0) \qquad (4)$$

where χ was the matrix of eigenvectors of the relaxation matrix **R**, λ was the diagonal matrix of eigenvalues, and τ_m was the mixing time (as defined in Chapter II).

Empirical correlation time reduction factors are incorporated during refinement to take into account differential motion:

$$\tau_{ij} = \tau_c * (S_i * S_j) \qquad (5)$$

As discussed in Chapter II, the product S_{ij}^2 relates to the order parameter, S^2 (Lipari & Szabo, 1982) and can vary between 0 and 1. Values of 0.65, 0.85 and 0.9 were empirically chosen for components of the order parameter to reflect the increased motion of all thymine methyl groups, sugar 2' and 2" methylene protons, and the 5' and 3' terminal residues, respectively, with an overall correlation time, τ_c, of 3.0 nanoseconds (Chapter II).

Structure refinement began with 100 steps of energy minimization using a dihedral force constant of 50 kJ•mol^{-1} rad^{-2} and an NOE force constant of 1000 kJ•mol^{-1} $(\Delta NOE)^{-2}$. A further 100 steps of minimization with CNOE set to 2000 kJ•mol^{-1} $(\Delta NOE)^{-2}$ completed the refinement procedure[1]. The residual NOE factor R was used to assess the fit of the NOEs calculated from the structure to observed NOE intensities (Chapter II):

$$R = \frac{\Sigma | NOE_{obs} - NOE_{calc} |}{\Sigma NOE_{obs}} \qquad (6)$$

where the summation runs over the number of observed NOE cross-peaks at mixing times of 200 and 500 milliseconds. Helical parameters of the final structures were analyzed with the programs AHELIX, BROLL, and CYLIN (Fratini et al., 1982; Dickerson et al., 1985).

---

[1]Larger NOE force constants cause unacceptable energies or distortions in the proton stereochemistry.

C. Results

1. Resonance Assignment

A prerequisite for the determination of a solution structure by NOE measurements is the assignment of resonances to specific protons of the macromolecule. Resonances in nucleic acids can be identified by a combination of COSY and NOESY experiments which provide, respectively, through-bond and through-space connectivities between individual nuclei (Aue et al., 1976; Nagayama et al., 1980; Jeener et al., 1979). COSY cross-peaks are observed for a pair of protons that interact through their NMR scalar coupling constant. The protons are generally two bonds apart, or three bonds apart with a favorable intervening dihedral angle. NOESY cross-peaks are observed between spatially proximate protons (generally with interproton distances of less than 5 Å). A NOESY spectrum is illustrated for GTAC in Figure III-2.

Figure III-3 shows the well resolved region of NOESY spectra of the two DNA octamers indicating possible close approaches of base H2, H6, and H8 protons (7 to 8.5 ppm) to cytosine H5 and sugar ring 1' protons (5.3 to 6.3 ppm). Adenine H8 and guanine H8 base protons generally resonate between 8 and 8.5 ppm, and between 7.5 and 8 ppm, respectively (Shindo et al., 1988). Pyrimidine H6 protons are between 7 and 7.7 ppm. COSY spectra of this spectral region each show only two strong peaks which correspond to the approximately 8 Hz three-bond coupling between cytosine base H5 and H6 protons. The short 2.5 Å length of this proton pair results in a strong NOE cross-peak at these resonance positions. The absence of strong cross-peaks (with an intensity similar to the cytosine H5-H6 pair) between guanine H8 and sugar 1' protons indicates no *syn* guanosine conformation, as is observed in left-handed DNA helices. NOE crosspeaks are observed from cytosine H5 to the base H8 protons preceding (5') in sequence (adenine for [d(G-T-A-C-G-T-A-C)]$_2$, guanine for [d(C-A-T-G-C-A-T-G)]$_2$), revealing that both duplexes are right-handed (Cohen, 1987). Moreover, CD spectra (taken by M. W. Germann & J. H. van de Sande, not shown) indicate right-handed helicity. Therefore, protocols developed for assignments of well resolved non-exchangeable protons in right-handed DNA apply (Feigon et al., 1982; Scheek et al., 1983; Hare et al., 1983; Chazin et al., 1986). Here, this assignment procedure is illustrated with the sequence [d(C-A-T-G-C-A-T-G)]$_2$, and the assignment method is updated to include backbone 4', 5' and 5" protons.

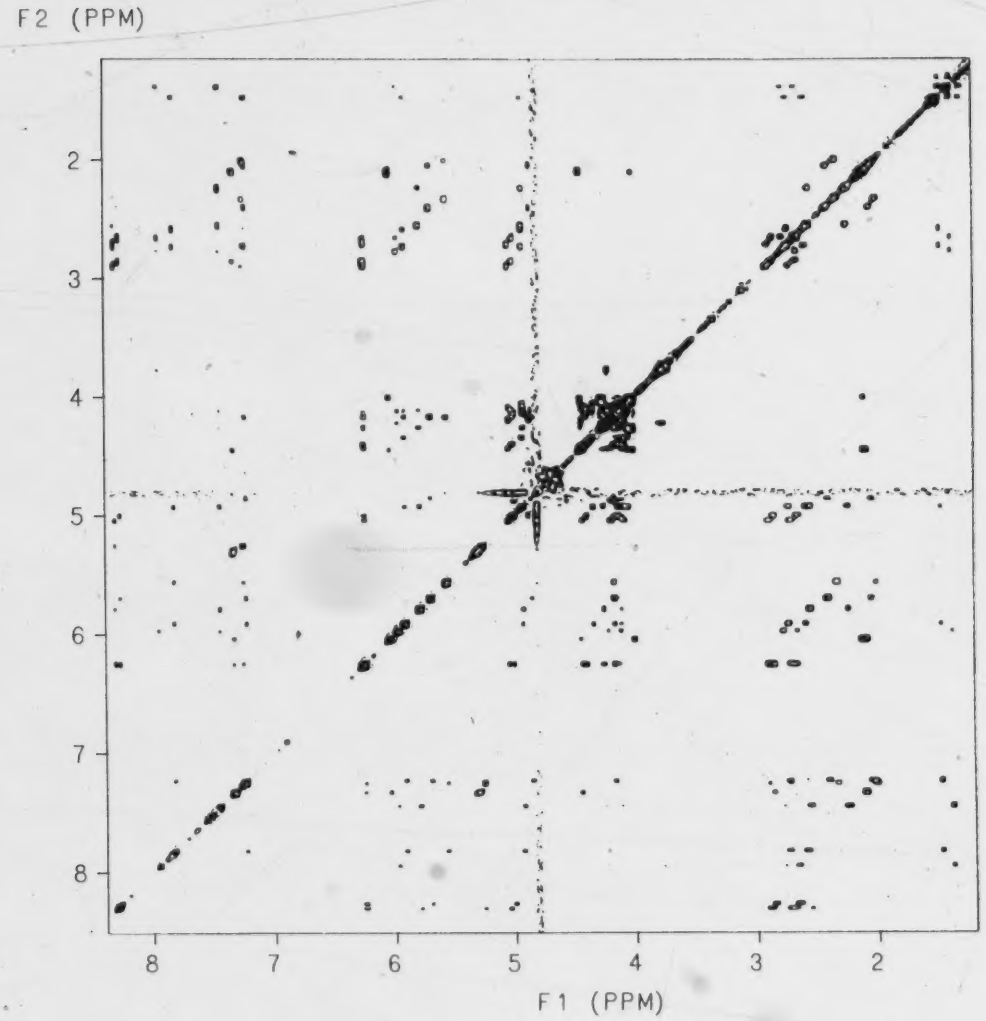


Figure III-2 NOESY spectrum of GTAC

The two-dimensional NOE (NOESY) ^1H NMR spectrum of GTAC in D_2O, 0.2 M KCl, pH 7, 20°C. A pair of protons close in space gives rise to a NOESY cross-peak between the corresponding proton NMR frequencies. Expansions of regions of this spectrum (and of NOESY spectrum for CATG) are shown in subsequent figures.

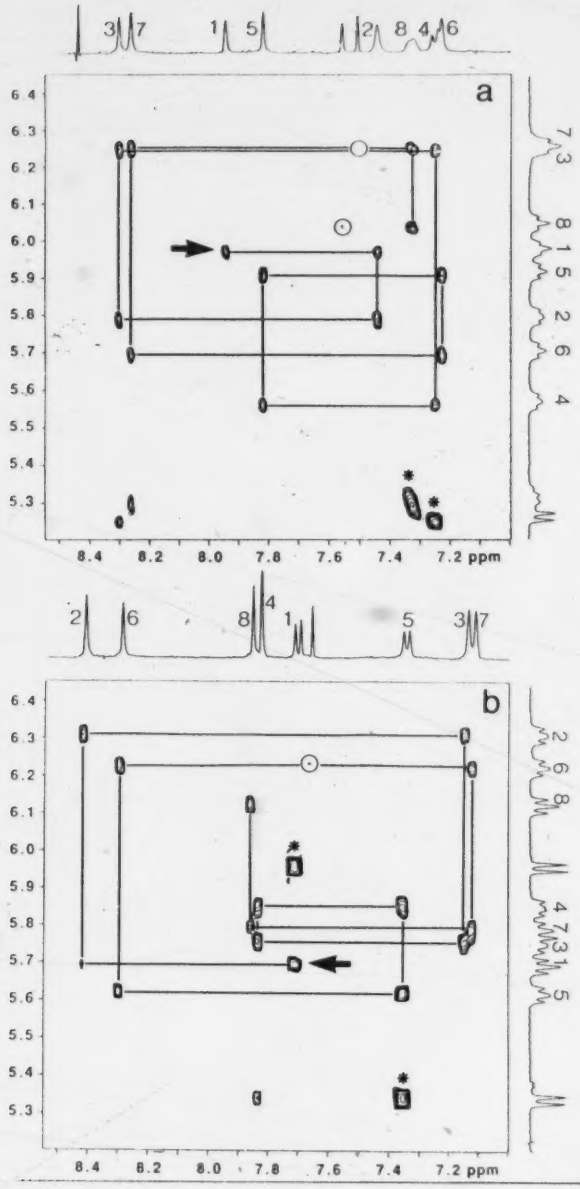


Figure III-3 400 MHz ^{1}H NMR NOESY spectra of the aromatic base and sugar 1' protons of DNA: (a) $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$; (b) $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$. The mixing time for the experiment is 500 milliseconds. Cross-peaks occur between the resonance frequencies for which corresponding protons are spatially proximate. Asterisks mark the cytosine H5-H6 NOE cross-peaks. Cross-peaks between adenine H2 protons and 1' protons are circled. For the first nucleotide of each duplex, the intra-residue base <-> 1' NOE cross-peak is indicated by an arrow. Residue numbers are given for H6/H8 protons and 1' protons on the one-dimensional spectra.

The thymidine H6 protons can be identified by a four-bond coupling constant resulting in COSY cross-peaks to their respective methyl resonances (near 1.3 ppm). The remaining adenine H2 resonances are attributed in one-dimensional spectra to the peaks with narrower linewidths and much longer spin-lattice relaxation times (T_1) than the other base protons, which is a consequence of the position of the H2 protons further into the middle of the DNA helix and resulting isolation from most other protons.

Differential line shapes for protons at either 5' or 3' ends of the DNA duplex reflect the dynamic behavior of these residues. In both helices, terminal 1' sugar protons are sharper than the others, indicating faster motion on a nanosecond time scale. In [d(G-T-A-C-G-T-A-C)]$_2$, one of the cytosine H6 base protons appears to be in slow chemical exchange ($t_{1/2} \geq 0.13$ seconds) with an alternate form. The H5-H6 NOE intensity and line width for this cytosine is similar to that for the other cytosine. The dynamics of this terminal residue is not investigated further, but must be kept in mind when interpreting structures that represent this nucleic acid.

Having attributed base protons to their residue type and knowing which sugar protons are likely at terminal positions, a sequence specific sequential assignment may now begin. For assignments in the base and 1' proton spectral region, advantage is taken of the fact that the protons of a DNA duplex in D_2O form a more or less equidistant linear array up and down each strand. Furthermore, in right-handed helices, base protons are near only two 1' sugar protons--their own and that of the preceding (5') residue, but not the 1' of the succeeding (3') residue (Figure III-4; Appendix 2).

In [d(C-A-T-G-C-A-T-G)]$_2$, one of the two cytosine H6 protons has a cross-peak to its own H5 proton, and to only one H1' proton. This represents the terminal cytosine H6 and its H1' since the 5' nucleotide is absent. The sharper H1' resonance supports the assignment. The H1' has the cross-peak to the C_1 H6, but also to an adenine H8, which is therefore the second residue. In turn, A_2 H8 has a cross-peak to its own intra-residue 1' proton, as well as the C_1 1' proton already noted. This connectivity pattern continues to the 3' terminal G_8 residue and is also observed throughout [d(G-T-A-C-G-T-A-C)]$_2$. A similar sequential procedure (Figure III-4) exists between base H8/H6 and H2' resonances and between base H8/H6 and the 2" protons, confirming assignments. Adenine H2 base proton assignments are made by the observation of weak cross-peaks to intra-residue 1' and succeeding 1' protons. Assignments into the ring can be made by examination of COSY spectra linking the assigned 1' protons to the 2' and 2" protons three bonds distant (Figure III-5). For all sugar ring conformations, the 1' proton is always closer, and has a larger NOE, to the 2" than to the 2' proton, thereby assigning the protons on the 2' carbon

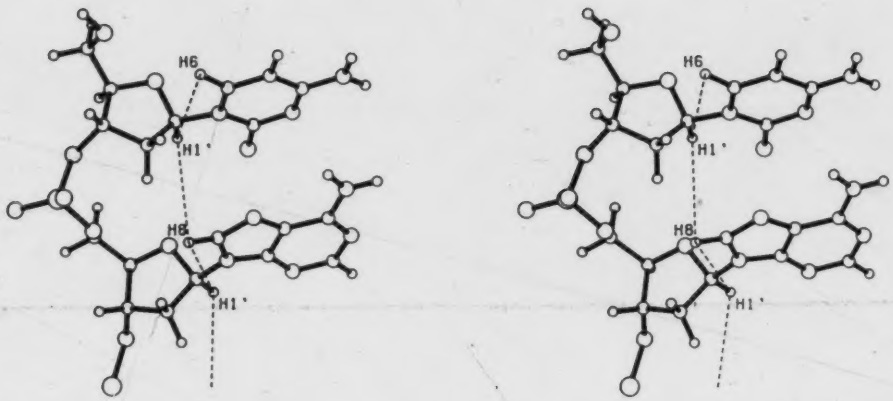


Figure III-4 Resonance assignment in nucleic acids

The stereo-diagram represents the first two nucleotides of the $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ duplex DNA. Spectral assignment most often begins with base H6/H8 and sugar 1' protons since they have the greatest spectral dispersion. In right-handed helices, the 1' proton is within 4 Å of the base proton of the same residue, and the base proton of the 5' residue (dashed line). A connectivity for a proton pair is represented by a cross-peak in the NOESY spectrum at the resonance frequencies of the 2 protons involved (Figure III-3).

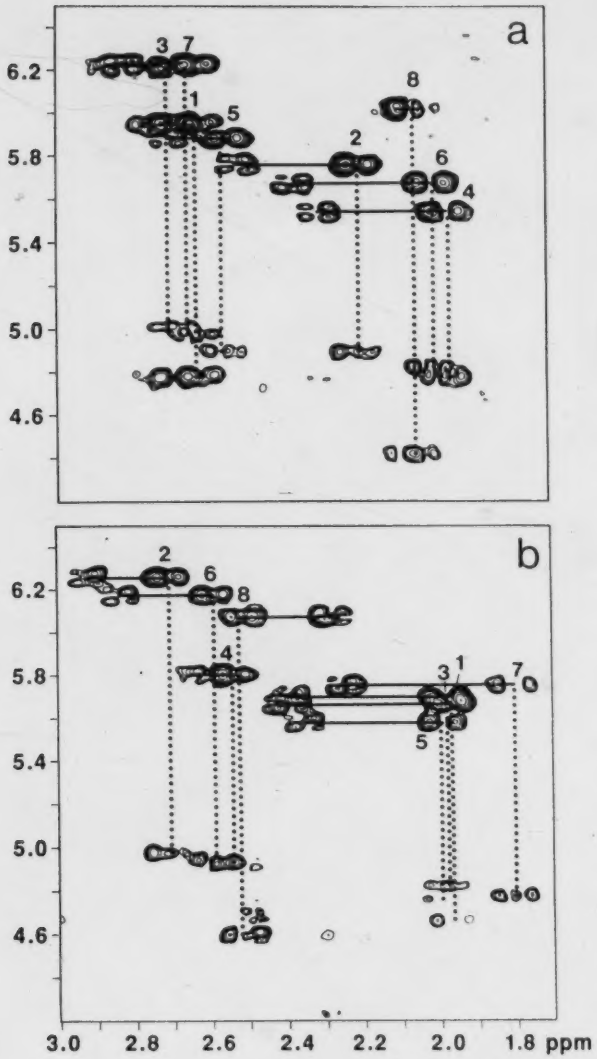


Figure III-5 Assignment of 2', 2", and 3' protons in 400 MHz COSY spectra of the DNA octamers: (a) [d(G-T-A-C-G-T-A-C)]$_2$; (b) [d(C-A-T-G-C-A-T-G)]$_2$. Cross-peaks in these spectra occur between protons three bonds apart which exist in a conformation such that there is a sizeable scalar coupling constant between them. In the top part of each figure, 1' protons have cross-peaks to 2' and 2" protons. There are corresponding cross-peaks for each residue between 2' and 3' protons (lower half).

stereospecifically. Excepting 3' termini, for any given residue the 2' proton is upfield of the 2" proton (Hare et al., 1983). 3' protons are assigned by COSY 2'<-> 3' cross-peaks. The absence of strong 2"<-> 3' cross-peaks (except for terminal residues) indicates that $J_{2"3'}$ is very small, and has implications for the sugar ring conformation (see below).

Assignment of the 4' and 5', 5" protons is illustrated for GTAC and CATG in Figure III-6. In right-handed DNA, these protons are nearest, and NOE intensities are calculated to be strongest, to the intra-residue base H6/H8 protons. Assignments are more conventionally made (Gronenborn & Clore, 1985) by examining COSY and NOESY spectral regions between the 3' and 4', 5', and 5" protons. COSY cross-peaks between 3' and 4' protons are repeated in the NOESY spectra, but with additional intra-residue 3'-5' and 3'-5" correlations (Figure III-7). The spectrum shown in Figure III-6 is more useful, as the effect of the residual HOD signal and limited spectral dispersion is removed from one frequency axis. 5' and 5" protons were not assigned stereospecifically because of the limited dispersion of chemical shifts in the 4', 5', 5" COSY and NOESY spectral regions. NOE intensities involving some 5' and 5" protons would also be modified by strong coupling effects (Kay et al., 1986) not taken into account in this study. Non-exchangeable proton assignments for GTAC and CATG are presented in Table III-1.

Several trends in chemical shifts reflect the alternating nature of the purine-pyrimidine sequences, and will be useful in future assignment of other related sequences. As has been noted before (Shindo et al., 1988), base H8 protons of purines resonate at lower field than H6 protons of pyrimidines. For alternating purine-pyrimidine sequences, generally all proton resonances, except the 5' and 5" protons, are at lower field for a purine nucleotide. This is a consequence of the chemical nature of the larger purine ring, and the conformational preference of six-member pyrimidine rings for a more negative χ torsion angles about the C1'-N bond. Base H6/H8 protons nearer the 5' end of the duplex are at higher frequency than those of the same type of residue four residues down the chain. The trend decreases further into the chain, and eventually reverses. The chemical shifts do not vary within 0.01 ppm between 5 and 30°C (except for C_8 H6 of $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$), indicating only limited fraying or reduced base stacking at the ends of the helices. These instead reflect chain termination effects, with the chemical shift being sensitive to both the chemical nature and conformational change induced by the lack of any additional nucleotides (Grütter et al., 1988). The terminal base-pairs are most affected, but some distortion may extend further into the duplex.

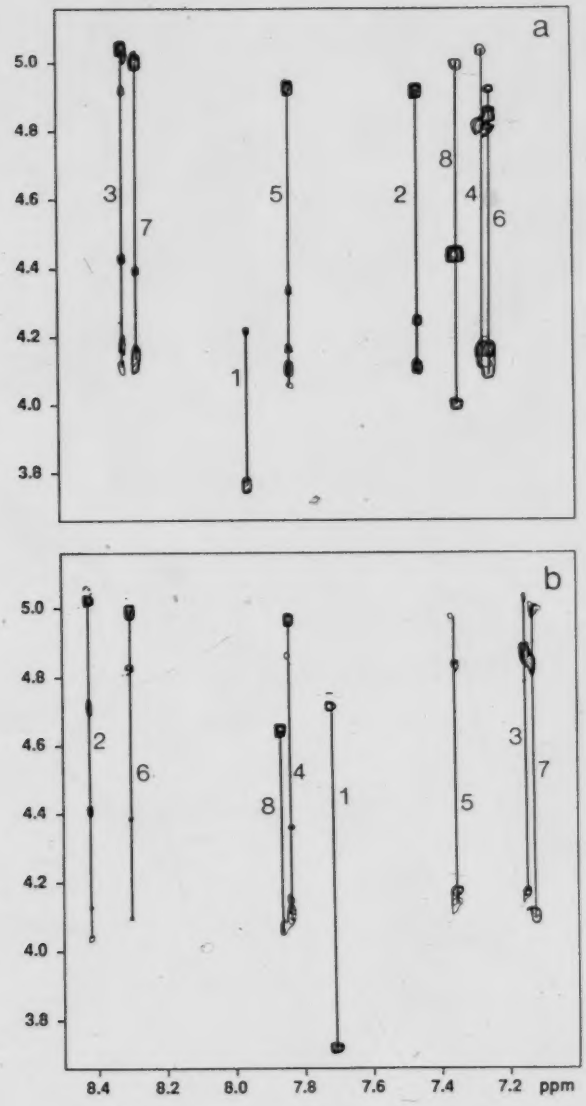


Figure III-6 Base H6/H8 NOE connectivities to backbone 3', 4', 5' and 5" protons (a) $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$; (b) $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$. The 500 millisecond NOESY contour plots show close approaches of pyrimidine H6 or purine H8 protons to intra-residue 3', 4', 5', and 5" protons.

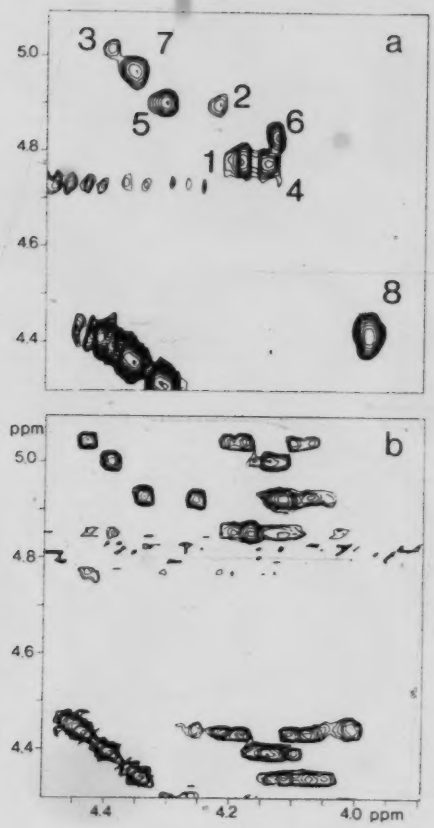


Figure III-7 Assignment of 3', 4', 5' and 5" protons for $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$
(a) COSY spectrum at 26°C (b) 500 millisecond mixing time NOESY spectrum at 20°C. Because the sugar ring pseudorotation angle is between 100 and 160°, there exists a dihedral angle giving rise to a coupling constant between the 3' proton and the 4' proton three bonds distant. Therefore, in the COSY spectrum, all 3'-4' protons are observed. The 3' and 4' protons are also near the 5' and 5" protons, which show as NOESY cross-peaks. Interpretation of the equivalent spectral regions in $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ is hindered by limited spectral dispersion of the 3' and 4' proton resonance frequencies, and by resonances of small small molecular weight impurities, which otherwise did not interfere. The NOESY spectrum shown in Figure III-6 was more useful, as one frequency axis is removed from the residual HOD.

Table 1 Proton chemical shift assignments for
[d(G-T-A-C-G-T-A-C)]$_2$ and [d(C-A-T-G-C-A-T-G)]$_2$

	Chemical shifts[a] (ppm)								
Residue	H6/H8	H2/H5/ CH_3	1'	2'	2"	3'	4'	5'	5"
G_1	7.94		5.98	2.68	2.77	4.80	4.22	3.78	3.78
T_2	7.44	1.40	5.80	2.26	2.58	4.93	4.26	4.14	4.03
A_3	8.31	7.52	6.25	2.73	2.93	5.05	4.45	4.09	4.19
C_4	7.25	5.27	5.57	2.03	2.36	4.79	4.18	4.23	4.10
G_5	7.82		5.92	2.58	2.76	4.94	4.32	4.12	4.07
T_6	7.22	1.49	5.71	2.08	2.42	4.85	4.17	4.11	4.22
A_7	8.26	7.57	6.25	2.67	2.86	5.00	4.39	4.15	4.09
C_8	7.35	5.35	6.05	2.16	2.08	4.47	4.04	4.49	4.25
C_1	7.72	5.98	5.72	2.00	2.43	4.72	4.06	3.72	3.72
A_2	8.42	7.82	6.32	2.78	2.95	5.03	4.42	4.03	4.14
T_3	7.16	1.42	5.77	2.03	2.42	4.87	4.19	4.27	4.08
G_4	7.82		5.87	2.62	2.[illegible]	4.97	4.36	4.15	4.08
C_5	7.37	5.36	5.63	2.07	2.38	4.84	4.17	4.13	4.10
A_6	8.29	7.68	6.23	2.64	2.87	5.00	4.38	4.16	4.06
T_7	7.12	1.46	5.82	1.85	2.28	4.82	4.11	4.24	4.08
G_8	7.87		6.13	2.58	2.36	4.65	4.16	4.04	4.07

[a]Chemical shifts are given relative to 2,2-dimethyl-2-silapentane-5-sulphonate at 20°C. 5' and 5" protons are not assigned stereospecifically.

2. Distance determination

Distances between protons are obtained using a distance extrapolation procedure (Chapter II). Distances are derived at each mixing time, plotted against the mixing time, and extrapolated back to zero mixing time as a first order correction for spin diffusion effects:

$$r_{ij} = \lim_{\tau_m \to 0} r_{ij}(\tau_m) = \lim_{\tau_m \to 0} \left\{ \frac{NOE_{ref}(\tau_m) * r_{ref}(\tau_m)^6}{NOE_{ij}(\tau_m)} \right\}^{\frac{1}{6}} \qquad (7)$$

For base and 1' sugar protons, the cytosine H5-H6 reference distance of 2.46 Å is used. Remaining distances between non-exchangeable protons are determined using the average $NOE_{ref}*r_{ref}^6$ products for the C2' methylene 2'<->2" proton pair (1.76 Å) and the 1'<->2" pair (2.3±0.1 Å for all sugar puckers). Upper and lower bounds on these distances are estimated from the distance extrapolation curve. For all distances greater than 3.5 Å and involving both an aromatic base and either of the C2' methylene protons, upper bounds are increased by 0.2 Å to account for systematic distance under-determination for this arrangement of protons (Chapter II). The alternating purine-pyrimidine DNA octamers are self-complementary so that restraints are entered in symmetry related pairs. Experimental restraints are given in detail in Appendices 3 and 4, and are summarized in Table III-2.

Table III-2 Numbers and types of experimental restraints used in structure determinations

	NOE intensities		Distances		Dihedral angles	
Structure[a]	Mix 200[b]	Mix 500[b]	NOE derived[c]	Watson-Crick[d]	Glycosidic	Right-handed helical restraints
GTAC	258	280	244	48	80	56
CATG	220	278	208	48	80	56

[a]GTAC and CATG refer to $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$, respectively. [b]Mix 200 and Mix 500 are the mixing times of the NOESY experiment, in milliseconds. These intensities are used for NOE-based refinement. [c]NOE derived distances are obtained from experimentally observed NOE intensities with mixing times of 100 to 500 milliseconds. [d]Watson-Crick base-pairing restraints.

3. Glycosidic dihedral angles

The geometry of the five-membered sugar ring of DNA can be described by five torsion angles ν_0 - ν_4. Because of ring closure, the values of ν_0 - ν_4 are interrelated:

$$\nu_n = \nu_{max} * \cos [P + 144* (n-2)] \qquad n=0,4 \qquad (8)$$

where ν_{max} is the maximum amplitude of the sugar ring pucker, and P is the sugar pseudorotation angle (Altona & Sundaralingam, 1972). The magnitude of three-bond coupling constants between protons is dependent on the intervening dihedral angle, and therefore reflects the pseudorotation angle that specifies the conformation of the sugar ring (Hosur et al., 1986).

Except for 3' termini, 2"<->3' correlations are absent in the COSY spectra (Figure III-5). This indicates that the sugar pseudorotation angle is between 100 and 250 degrees for all non-(3')terminal sugar rings (Hosur et al., 1986). The observation of 3'<->4' correlations (not shown) narrows the allowed range of the sugar pseudorotation angle to be between 105 and 175°. Individual $J_{1'2'}$ and $J_{1'2"}$ coupling constants measured from one-dimensional spectra indicate residue-to-residue variation and further restrict the pseudorotation angle (Chary et al., 1988; Hosur et al., 1988; Rinkel & Altona, 1987). Allowed ranges of glycosidic dihedral angles are obtained for each nucleotide from the coupling constant data by assuming a ν_{max} of 40° (Rinkel & Altona, 1987), and using equation 8.

4. Right-handed DNA helix restraints

To preserve the right-handed character of the DNA during the molecular dynamics calculations, it was sometimes necessary to constrain backbone dihedral angles to be in a broad allowed region of torsional angle space (Gronenborn & Clore, 1989). The allowed angles (α, -90 to -30°; β, < -145° and >135°; ε, < -60° and >140°; ζ, < -45° and > 150°) are derived from a table of conformation angles found in the different DNA types (Suzuki et al., 1986) and from considering individual variations in right-handed helices from X-ray crystallographic studies (Dickerson et al., 1985). These right-handed helix restraints would cause no violations for any of the average A, B, alternating-B, C, D, or wrinkled D DNA forms, nor for any individual angles found in the best studied single crystal X-ray structures of B DNA (Dickerson & Drew, 1981; Privé et al., 1987), except for two β angles (of residues G_3 and C_{12}) in the B DNA form of a phosphothioate analogue of DNA

(Cruse et al., 1986).

During molecular dynamics runs, base-pairs were kept Watson-Crick hydrogen-bonded by distance restraints between bases. These were: for all base-pairs: r(C1' - C1', across the base-pair) = 10.87±0.20 Å; for AT base pairs: r(A_{N6} - T_{O4}) = 2.8±0.1 Å, r(A_{HN} - T_{O4}) = 1.7±0.1 Å, r(A_{N1} - T_{H3}) = 1.7±0.1 Å, r(A_{N1} - T_{N3}) = 2.8±0.1 Å; and for GC pairs: r(G_{O6} - C_{HN}) = 1.6±0.1 Å, r(G_{O6} - C_{N4}) = 2.7±0.1 Å, r(G_{N1} - C_{N3}) = 2.8±0.1 Å, r(G_{H1} - C_{N3}) = 1.7±0.1 Å, r(G_{N2} - C_{O2}) = 2.8±0.1 Å, r(G_{HN} - C_{O2}) = 1.7±0.1 Å (Arnott & Hukins, 1972).

5. Molecular dynamics calculations and NOE-based refinement

The A and B DNA starting structures for each MD run have a rms deviation of 4.3 Å (Table III-3). For each duplex, the application of molecular dynamics including the experimental restraints results in convergence to structures with an atomic rms deviation of 0.65 Å and 0.66 Å, for GTAC and CATG, respectively. This is comparable to the rms fluctuations over the last five picoseconds of the molecular dynamics run (0.50 and 0.62 Å, respectively). The MD structures are shown in Figures III-8(a) and (b). The structures drawn in bold are results starting from A DNA models; the others have initial B DNA models. The two structures for each duplex appear to be essentially identical, notwithstanding motional dynamics. It is therefore appropriate to average over the last 5 picoseconds of both runs starting from A and from B DNA (i.e. averaging over 10 picoseconds of MD for each DNA octamer duplex). The rms fluctuations increase to only 0.55 and 0.65 Å for GTAC and CATG, respectively.

After non-exchangeable protons are added in geometrically reasonable positions, these averaged structures are each subjected to refinement directly based on the NOE intensities, and not on the derived, approximate distances. Initial NOE R factors are 0.24 and 0.26 respectively for GTAC and CATG. Energy minimization reduces the NOE residual to 0.19 and 0.23, respectively. The larger R factor for CATG is likely not due to a poorer structural determination *per se*, but reflects the lower concentration of this duplex, with concomitantly lower signal-to-noise ratio in NOESY spectra, and increased error in the NOE intensities. In Chapter II it was shown that an R factor of less than approximately 0.20 indicates that the structure is in agreement with the NOE data within a reasonable experimental error. The refinement procedure results in little change from the average MD structure (less than 0.1 Å rms deviation), which is consistent with structures that are energetically near a global minimum. Long NOE-restrained molecular dynamics simulations, which would sample much more conformational space than NOE-restrained energy minimization, could not be

Table III-3 Atomic rms differences (Å)
between alternating pyrimidine-purine DNA structures

	Overall rms difference (Å) for $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$				
	A-DNA	B-DNA	MD-A	MD-B	R_{ave}
A-DNA	-	4.3	3.9	3.4	3.7
B-DNA	4.3	-	1.13	1.5	1.3
MD-A	4.3	1.2	-	0.65	0.33
MD-B	3.9	1.4	0.66	-	0.34
R_{ave}	4.1	1.2	0.34	0.34	-
	Overall rms difference (Å) for $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$				

A-DNA and B-DNA are starting structures with regular A and B geometries, respectively. MD-A and MD-B are average restrained molecular dynamics structures. NOE-based refinement results in the R_{ave} structures. Numbers above the diagonal indicate the rms atomic deviation between two structures for $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$. Numbers below the diagonal show the comparison for $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$.

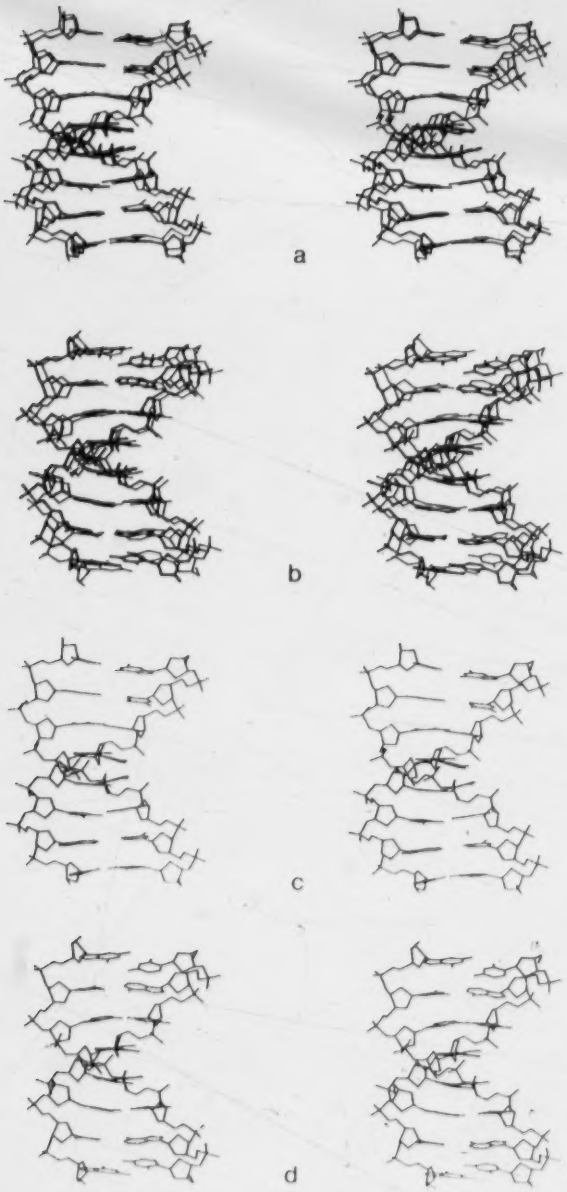


Figure III-8 Structures of alternating purine-pyrimidine DNA
(a) [d(G-T-A-C-G-T-A-C)]$_2$ restrained MD structures; (b) [d(C-A-T-G-C-A-T-G)]$_2$ MD structures; (c) NOE-refined structure of [d(G-T-A-C-G-T-A-C)]$_2$; (d), NOE-refined structure of [d(C-A-T-G-C-A-T-G)]$_2$. Protons have been omitted to preserve clarity in the diagrams.

undertaken because of the computer time required (Chapter II). The final structures are energetically reasonable, showing lower energies than the average MD structure (Table III-4).[2] The refined structures are shown in Figures III-8(c) and (d). Helical parameters are obtained by fitting the best overall helix, excluding the terminal base-pairs. GTAC has an average helix rotation of 34.8°, or 10.4 residues per turn, and a mean residue-to-residue rise of 3.27 Å. CATG has an average helix rotation of 34.4°, or 10.5 residues per turn, with a mean rise of 3.34 Å. Both structures have atomic rms deviations from classical B DNA of approximately 1.25 Å.

Convergence from the widely different starting structures indicate that conformational space has been well sampled and that final refined structures provide reasonable representations of the structure of alternating purine-pyrimidine sequences in solution (Nilges et al., 1987; Chapter IV). Despite the fact that NOEs give the spatial relationships between protons only a few Ångstroms apart, and coupling constants indicate torsion angles for protons separated by three bonds, the structures shown in Figure III-8 have global features similar to known oligonucleotide structures determined crystallographically. The short range nature of the NMR information suggests that local conformational parameters should be better determined, which are therefore discussed in some detail.

---

[2] Although energies for all MD and final structures are approximately the same, energy-minimized starting A and B DNA structures are not. This is a consequence that the first 200 steps of energy minimization are insufficient to relieve steric clashes present in the initial models for some structures (e.g. CATG B DNA_m).

Table III-4 Energy terms for alternating purine-pyrimidine structures

	Potential energy (kJ • mol^{-1})									
GTAC	Total	elec.	vdW	bond	angle	imp. dihedral	torsion	distance restraint	dihedral restraint	NOE restraint
A DNA_m	-797	-261	-1513	10	166	27	561	148	66	-
B DNA_m	-808	-259	-1427	12	189	16	631	32	2	-
MD_{ave}	-829	-268	-1492	22	211	104	632	85	3	-
R_{ave}	-944	-261	-1561	10	184	103	630	-	2	70
CATG										
A DNA_m	-529	-567	-1055	40	178	25	585	214	50	-
B DNA_m	-276	-568	- 668	35	207	6	667	45	1	-
MD_{ave}	-865	-262	-1484	16	200	109	602	71	1	-
R_{ave}	-890	-257	-1530	10	193	105	647	-	1	110

GTAC and CATG refer to $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$, respectively. A DNA_m and B DNA_m are energy-minimized starting structures with regular A and B geometries, respectively. MD_{ave} refers to the average restrained molecular dynamics structure resulting from averaging over the last 5 picoseconds of both MD runs starting from A and B DNA, the addition of non-exchangeable protons, and energy minimization to correct distortions caused by the averaging procedure. R_{ave} is the final refined structure after NOE-based refinement of MD_{ave}. Elec. refers to the electrostatic energy, vdW to the van der Waals energy, and imp. to the improper or forced dihedral angle energy. Energies are given for initial A and B DNA models after energy minimization with distance restraints, with CDIS equal to 500 kJ•mol^{-1} nm.$^{-2}$ and CDLR at 5 kJ•mol^{-1} rad^{-2}. For the remaining structures, CDIS and CDLR are 20 and 10 times larger, respectively (10000 kJ•mol^{-1} nm.$^{-2}$ and 50 kJ•mol^{-1} rad^{-2}). The rms distance restraint violations are approximately 0.50 and 0.25 Å for the A and B DNA_m structures, and 0.075 Å for the MD_m structures. A NOE restraint potential replaces the distance restraint potential for refined structures. For R_{ave} and MD_m structures, tabulations of total energy exclude the contributions from non-exchangeable protons.

D. Discussion

1. Structural details of alternating purine-pyrimidine DNA octamers

Structural parameters typical for describing nucleic acid structure (Dickerson et al., 1985) are shown in Figure III-9. The top four parameters, helix twist, roll, δ1-2, and propeller twist, represent conformational features shown to be most base-sequence dependent (Dickerson, 1983), and a set of rules has been proposed to explain their behavior (Calladine, 1982) based on a steric clash model. In nucleic acids, a small propeller twist enables the adjacent bases along one strand to overlap, and stack more efficiently. Along one strand, bases rotate in the same sense with no interference. However, purines extend beyond the center of the base-pair, and, because of the anti-parallel nature of DNA, bases of the opposite strand have their propeller rotation in the opposite direction with respect to the first. This results in steric clashes occurring between the purines of adjacent base-pairs. The unacceptably short cross-chain contacts are more severe in the minor groove at pyrimidine-purine steps than in the major groove at purine-pyrimidine steps. The steric clash can be relieved by one or more of four strategies: (1), by decreasing the local helix twist angle between adjacent base-pairs; (2), by increasing the roll angle between base-pairs; (3), by sliding one or both base-pairs along its long axis, pulling the purine out of the base-pair stack, leading to a larger purine δ torsion angle and decreased pyrimidine δ torsion angle; and (4), by flattening the propeller twist in one or both base-pairs.

These rules (results not shown) predict an out-of-phase alternating behavior in the helix twist, roll, and difference in δ torsion angles across the base-pair, and predict an in-phase behavior for the propeller twist. This is as observed, except for the propeller twist angle of the terminal base-pairs and for the helix twist of the middle base-step in [d(C-A-T-G-C-A-T-G)]$_2$.

The distance between adjacent purine C2 atoms in the minor groove averages 3.6 Å (range 3.4 to 3.7 Å) for the two octamers. This compares well to the crystal structure of the B DNA dodecamer, [d(C-G-C-G-A-A-T-T-C-G-C-G)]$_2$ (Dickerson & Drew, 1981), where this distance also averages 3.6 Å (range 3.5 to 3.7 Å). Of the mechanisms for relieving steric clash, base roll seems most important. For example, the central base step of [d(G-T-A-C-G-T-A-C)]$_2$ is suggested by Calladine's (1982) rules to relieve the clash by decreasing the twist angle. In fact, the opposite occurs, and the clash is relieved by a large 12° base-pair roll. Calladine's rules (1982) appear to predict only the roll angle well. The

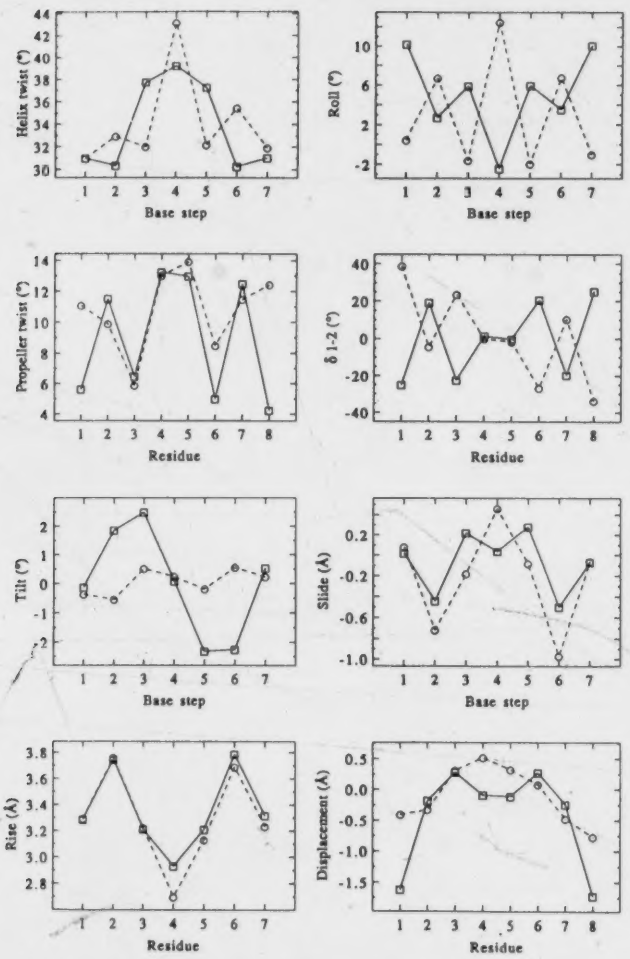


Figure III-9 Conformational parameters for alternating purine-pyrimidine DNA sequences $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ (--o--), and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$ (—□ —). Helical twist: Twist angle between adjacent mean base-pairs. Roll: Angle between the normals of adjacent base-pairs in the direction of the minor groove. Propellor twist: Angle between normals to the base planes of a pair, viewed along the long axis of the base-pair. δ1-2: Difference in δ (C4'-C3') torsion angles for the residue at each end of a base-pair. Tilt: Angle between the normals of adjacent base-pairs in the direction of the long axis of the base-pair. Slide: The relative displacement of two successive base-pairs along the direction of their long axis. Rise: Distance between mean adjacent base-pairs. Displacement: Distance between the base-pair midpoint and the helix axis. Precise descriptions for the conformational parameters are given by Dickerson et al. (1985).

generally poor correlations for the other parameters has been observed in a number of studies, suggesting that the Calladine steric clash model is insufficient to account for all local helix perturbations (Dickerson et al., 1985; Scalfi Happ et al., 1988; Nilges et al., 1987; Lefèvre et al., 1987; Privé et al., 1987; Chapter IV).

The arrangement between adjacent base-pairs optimizes base-stacking interactions by allowing for intrastrand purine-purine stacking at purine-pyrimidine steps and some interstrand purine-purine stacking at pyrimidine-purine steps (Figures III-10 and III-11). Interstrand base stacking is reminiscent of A DNA structure, and more so of a special alternating-B DNA model structure proposed by Klug et al. (1979), and has been observed before in alternating purine-pyrimidine sequences (Nilges et al., 1987). The rise between base-pairs at the center of each of the octamers decreases to 2.8 Å, which is more characteristic of A DNA. The distance from the C6-C8 vector of the base-pair to the helix axis, the displacement, averages +4 Å for A DNA and -0.2 Å for B DNA. Although certainly not large in magnitude here (Figure III-9), slightly more positive values of displacement for the central core of the octamers reflect part of its A DNA type character. Greater base-stacking for purine-pyrimidine than pyrimidine-purine steps is observed in the crystal structure of a DNA dodecamer (Dickerson & Drew, 1981), and is predicted by fibre diffraction studies (Arnott et al., 1969). In our alternating purine-pyrimidine sequences, this occurs to an even greater degree. As was predicted in an alternating-B DNA model, the five-membered ring of the 5' purine is brought more directly over the C5 atom of the 3' thymine and its methyl group. This van der Waals interaction is proposed to make the structure particularly stable (Klug et al., 1979). Probably as a result of optimized stacking, there is somewhat less symmetry across the helix axis than is observed in the study of a DNA decamer of non-alternating sequence (Chapter IV).

Some of the other conformational parameters of Figure III-9 are independent of base sequence, and more dependent on the distance from the ends of the duplex and therefore may be influenced more by the effects of chain termination. The same behavior is observed for some of the torsion angles (Figure III-12). Although the sequences reverse, values of the α, β, and ε angles are generally parallel to each other in GTAC and in CATG. There is greater diversity in each of the angles than seen in the DNA decamer using the same methodology (Chapter IV), but not as much as is observed in single crystal B DNA studies (Dickerson et al., 1985; Cruse et al., 1986; Dickerson & Drew, 1981; Privé et al., 1987), which may be due to conformational averaging in solution. Backbone α, β, γ, ε, and ζ angles are all *gauche*$^-$, *trans*, *gauche*$^+$, *trans*, and *gauche*$^-$, respectively, and average torsion angles of the alternating purine-pyrimidine octamers correspond closely to the averages of

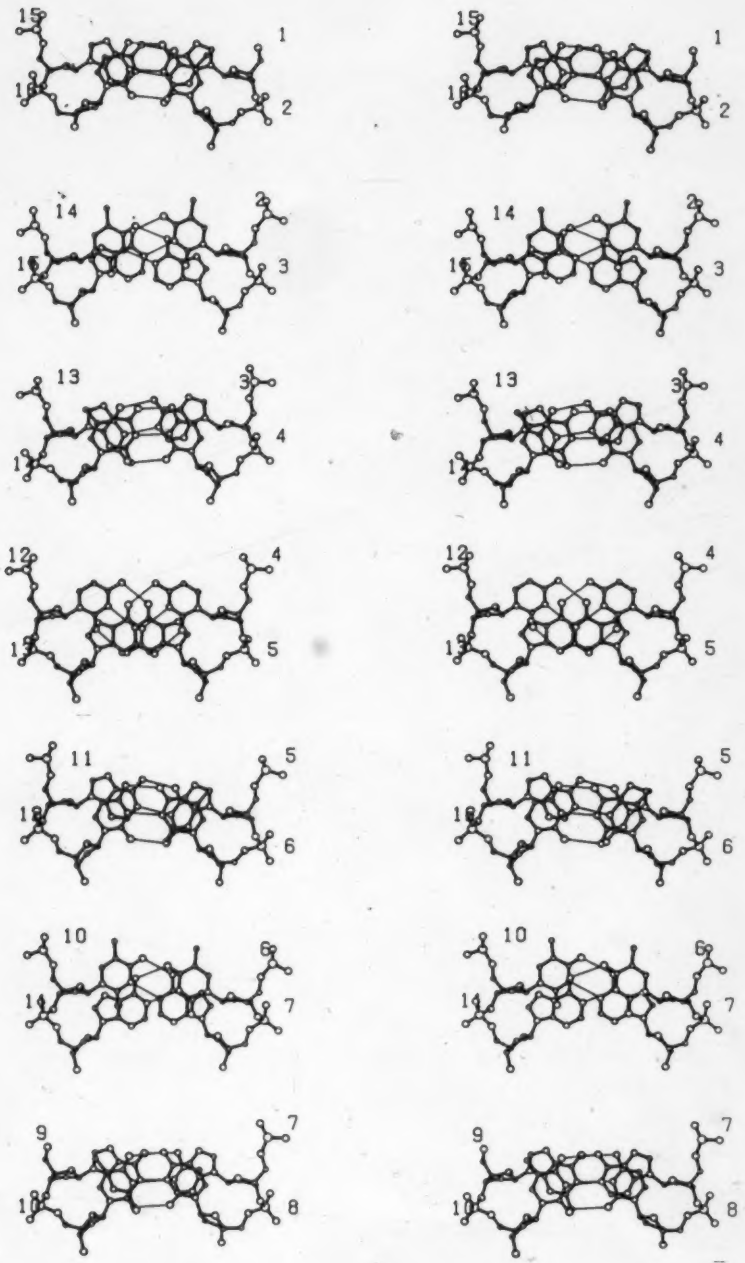


Figure III-10 Base-stacking arrangements in [d(G-T-A-C-G-T-A-C)]$_2$
Views down the helix axis are presented two base steps at a time.

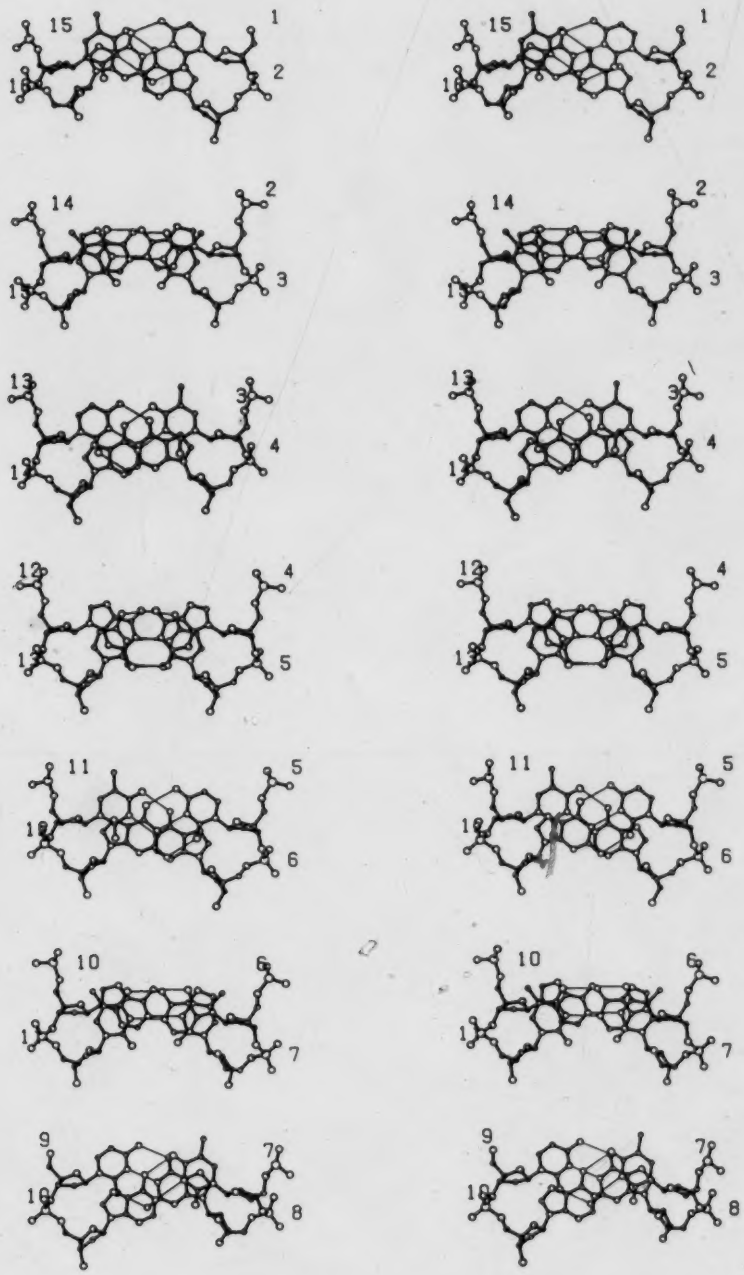


Figure III-11 Base-stacking arrangements in [d(C-A-T-G-C-A-T-G)]$_2$.

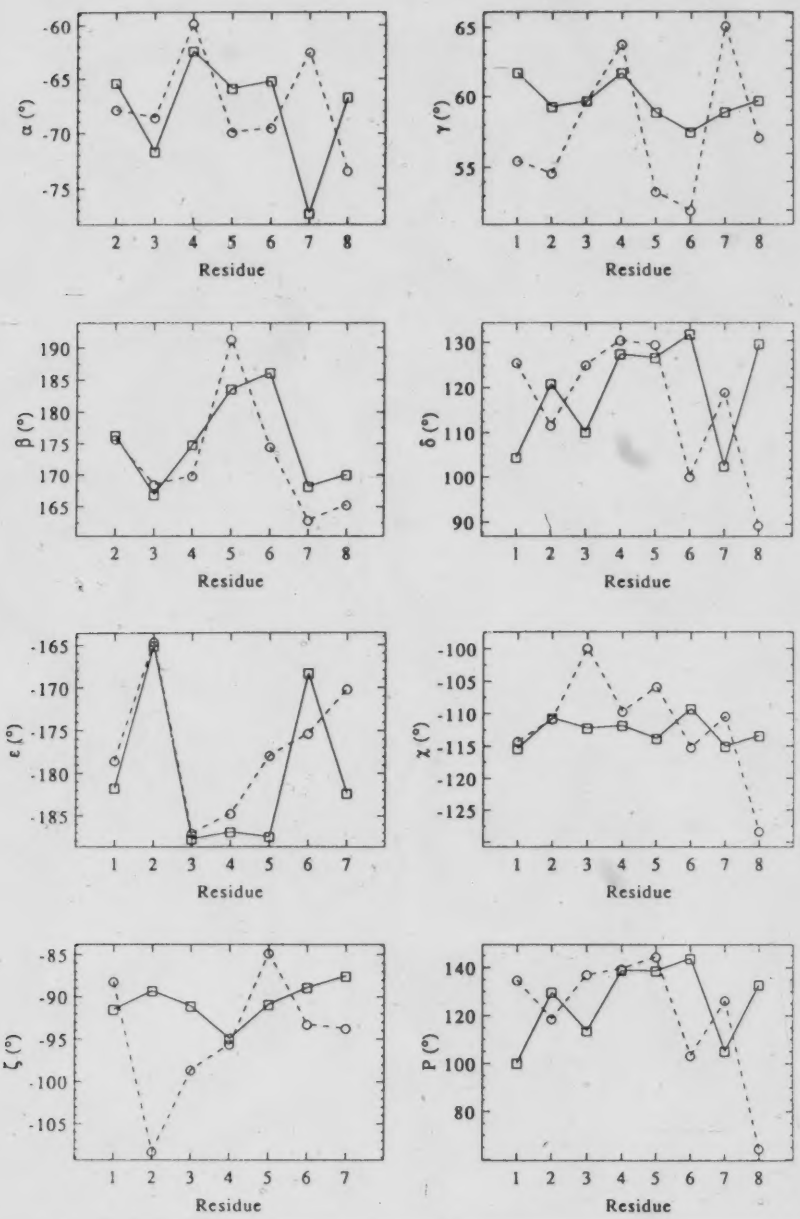


Figure III-12 Variation in the backbone and glycosidic bond torsion angles as well as the phase angles describing the sugar pucker for [d(G-T-A-C-G-T-A-C)]$_2$ (--o--) and [d(C-A-T-G-C-A-T-G)]$_2$ (— □ —). Main chain: P-$\overset{\alpha}{\cdots}$-O5'-$\overset{\beta}{\cdots}$-C5'-$\overset{\gamma}{\cdots}$-C4'-$\overset{\delta}{\cdots}$-C3'-$\overset{\varepsilon}{\cdots}$-O3'-$\overset{\zeta}{\cdots}$-P
Purines: O4'-C1'-$\overset{\chi}{\cdots}$-N9-C4 Pyrimidines: O4'-C1'-$\overset{\chi}{\cdots}$-N1-C2

single crystal B DNA structures (Dickerson & Drew, 1981; Chapter IV). No alternation between B_I and B_{II} (ε,ζ torsion angles that are *trans,gauche*$^-$ and *gauche*$^-$*,trans*, respectively) about the phosphates is observed for the alternating purine-pyrimidine octamers, unlike that observed in the crystal structure of an alternating cytosine-guanine B DNA phosphothioate deoxyoligonucleotide (Cruse et al., 1986). It is not surprising that the B_I conformation should be the one exclusively preferred in solution, as the B_I conformation is more common in crystallographic studies, especially in regions away from inter-molecular hydrogen bonds and crystal contacts (Dickerson et al., 1987; Chapter IV). Sugar rings are centered on a C1' *exo* to O1' *endo* range of pseudorotation angle (P is between 100 and 145°), except the poorly determined 3' terminal residues of $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$, which are C4' *exo* (Altona & Sundaralingam, 1972). No attempt was made to interpret our results for conformational averaging between C2' *endo* and C3' *endo*, or any other two possible sugar ring conformations. Purines do not display any tendency to 3' *endo* conformations, as seen in the alternating-B DNA model and in Z form DNA conformation. The χ, δ, and P angles do show some correlation with each other (Figure III-12), despite the angles falling into a narrow range. They best reflect the alternating nature of the purine-pyrimidine sequences.

2. The biological role of alternating purine-pyrimidine sequences

The key to understanding the properties of alternating purine-pyrimidine sequences is in the base stacking diagrams of Figures III-10 and III-11. 5' Purines stack much more efficiently over 3' pyrimidines than do 5' pyrimidines over 3' purines. This suggests dinucleotide purine-pyrimidine structural units. In the absence of external forces, such as crystal packing, super-helical stress, or binding by proteins, the DNA conformation is rather regular, without large alternating conformational characteristics to the backbone. However, the decreased stacking of pyrimidines over purines allows greater conformational flexibility at pyrimidine-purine steps, and leads to inducible phosphodiester linkage differences at these steps (Klug et al., 1979).

For example, when $[GC]_n$ and $[AT]_n$ copolymers are digested by DNAase I, most of the product oligonucleotides have a pyrimidine at their 5' ends, so that successive oligomers differ in length by two nucleotides (Lomonossoff et al., 1981). Presumably, the alternating purine-pyrimidine DNA octamers would display the same behavior. The structure determinations here suggest that upon binding the enzyme, dinucleotide units become more distinct, and provide a mechanism for distinguishing between the 5' and 3' directions about a pyrimidine residue (Klug et al., 1979; Cruse et al., 1986). The recent

crystallographic determination of a nicked DNA complex with DNAase I shows a widening of the minor groove resulting from protein residues filling this groove and their tight interaction with DNA (Suck et al., 1988). Although no local parameter was invoked in this study to explain the variation in cutting rates of phosphodiester bonds, both helical twist and roll angles have been implicated (Lomonossoff et al., 1981). Large positive roll angles for pyrimidine-purine steps (Figure III-9), in which the angle between the planes of the base-pairs opens towards the minor groove, might be the feature of alternating sequences that leads to preferred cutting by DNAase I at the 5' side of pyrimidine residues (Lomonossoff et al., 1981).

The dinucleotide structural unit is also an important feature in the formation of left-handed Z DNA tracts within B DNA. The major difference between B and Z DNA is related to the orientation of the base pair relative to the sugar phosphate chain (Wang et al., 1979). Conversion to Z DNA may be associated with an initial separation of the base pairs, and a rotation followed by a rejoining of the base pairs. The purine rotates about the glycosidic bond (torsion angle χ), resulting in the *syn* conformation. The entire pyrimidine rotates, base and sugar, retaining the *anti* χ conformation, and producing the characteristic zig-zag of the sugar phosphate backbone. The structures presented here imply that the initial separation event is at the pyrimidine-purine step.

3. Implications for structure determination of macromolecules in solution using NMR

This paper illustrates a reasonable approach to determine the structure of DNA duplexes to atomic resolution on submillimolar concentrations in solution. The incorporation of NOE-based refinement circumvents inaccuracies in distances caused by neglecting indirect cross-relaxation pathways for the NOE between two protons and therefore one of the most serious problems in structure determination using NMR techniques. Often at long mixing times, NOE cross-peaks can be observed between protons even 6 and 7 Å apart, but this information is lost since it cannot be directly related to a distance. The NOE refinement method indicates a way to obtain more parameters that cover a greater range of conformational space, and to better determine structures in solution by nuclear magnetic resonance.

Controversy exists over the extent of perturbation of structure by the approximate nature of force fields in restrained molecular dynamics calculations. The most poorly understood term in the force field is the electrostatic calculation of the non-bonded interactions. A limited study has been undertaken on the effect of the dielectric constant on nucleic acid conformation, in the presence of distance restraints. Structure determinations

were carried out on $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$, starting from an A DNA model, and averaging over the last 5 picoseconds of MD, in the same manner as above. The dielectric constant was varied from being equal to 20r (where r is the separation of the charged groups, in Å), to being equal to 1 (with no distance dependence). The electrical component to the total potential energy varied from -53 to -2623 $kJ \cdot mol^{-1}$ (cf. the MD_{ave} structure of Table III-4). The surprising result was that all structures met the distance restraints equally well ($E_{DIS} = 80 \pm 5$ $kJ \cdot mol^{-1}$), and had an average rms deviation from each other of about 0.5 Å, and had essentially identical conformational parameters. As has been noted before, the interproton distances are a direct measure of the actual solution structure under the experimental conditions employed, and are the principal driving force in determining the conformations of structures (Nilges et al., 1987).

There are, however, a number of limitations on the accuracy of the structure determinations presented here. Especially in the absence of a chosen molecular mechanics force field, it is still insufficiently understood to what extent the few hundred NOE intensities (or short distances) and dihedral angles define the three dimensional structure of oligonucleotides, including a description of dynamics (van de Ven & Hilbers, 1988). For example, there exists little direct information on backbone torsion angles, although the positioning of the nucleotide units relative to each other in the presence of even an approximate force field may be sufficient to localize the angles to a relatively narrow region of conformational space (Nilges et al., 1987).

As discussed in Chapter II, our initial motion model assumed that all interproton vectors tumble isotropically and are of invariant length. Therefore, besides some anisotropic behavior of motion, the NOE between protons not fixed by covalent geometry can be modified by distance fluctuations. The observed cross-peak is an approximate r^{-6} weighted average, and is therefore dominated by the close approach of the atoms. A more sophisticated model could be used, but the incorporation of model-free empirical parameters here is sufficient with respect to the precision of the NOE data experimentally collected. Although preferable to use the direct comparison between observed and calculated NOEs in a long restrained MD run (since a better search of conformational space would be performed and the step of making rather approximate distance determinations eliminated), the computational time would be cumbersome. Moreover, it is unclear whether the structure produced by using the NOEs directly in MD runs would be essentially more correct than that produced using approximate distances, followed by refinement with NOEs, given any inadequacies in the motional model used here (Chapter II).

E. References

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem Soc. 94*, 8205-8212.

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504-1509.

Arnott, S., Dover, S. D., & Wonacott, A. J. (1969) *Acta Crystallogr. sect. B 25*, 2192-2206.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229-2246.

Baleja, J. D., Moult, J., & Sykes. B. D. (1990a) *J. Magn. Reson.* (in press).

Baleja, J. D., Pon, R. T., & Sykes. B. D. (1990b) *Biochemistry* (in press).

Beaucage, S. L., & Caruthers, M. H. (1981) *Tetrahedron Letters 22*, 1859-1862.

Behling, R. W., Rao, S. N., Kollman, P., & Kearns, D. R. (1987) *Biochemistry 26*, 4674-4681.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684-3690.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187-217.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343-352.

Chaconas, G., & van de Sande, J. H. (1980) *Meth. Enzymol. 65*, 75-85.

Chary, K. V. R., Hosur, R. V., Govil, G., Chen, C., & Miles, H. T. (1988) *Biochemistry 27*, 3858-3867.

Chazin, W. T., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin, W. (1986) *J. Mol. Biol. 190*, 439-453.

Cohen, J. S. (1987) *Trends Biochem. Sci. 12*, 133-135.

Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F., & Kennard, O. (1986) *J. Mol. Biol. 192*, 891-905.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Israel J. Chem. 27*, 181-188.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419-441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761-786.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) *J. Biomol. Struct. Dynam. 5*, 557-579.

Dickerson, R. E., Kopka, M. L., & Pjura, P. (1985) in *Biological Macromolecules and Assemblies, Volume 2, Nucleic Acids and Interactive Proteins*, (Jurnak, F. A., & McPherson, A., Eds.) pp 37-126, 490-493, Wiley-Interscience, New York.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1985) *Science 230*, 82-84.

Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540-5541.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686-14707.

Germann, M. W., Schoenwaelder, K.-H., & van de Sande, J. H. (1985) *Biochemistry 24*, 5698-5702.

Gronenborn, A. M., & Clore, G. M. (1985) *Progr. Nucl. Magn. Reson. Spect. 17*, 1-33.

Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28*, 5978-5984.

Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Biophys. J. 16*, 279-286.

Hagen, F. K., Zarling, D. A., & Jovin, T. M. (1985) *EMBO J. 4*, 837-844.

Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319-336.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281-294.

Hosur, R. V., Govil. G., & Miles, H. T. (1988) *Magn. Reson. Chem. 26*, 927-944.

Hosur, R. V., Ravikumar, M., Chary, K. V. R., Sheth, A., Govil, G., Zu-kun, T., & Miles, H. T. (1986) *FEBS Lett. 205*, 71-76.

Jeener, J., Meier, B. H., Backmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546-4553.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179-182.

Kay, L. E., Scarsdale, J. N., Hare, D. R., & Prestegard, J. H. (1986) *J. Magn. Reson. 68*, 515-525.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Zhakked, Z., & Steitz, T. A. (1979) *J. Mol. Biol. 131*, 669-680.

Lefèvre, J.-F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076-5090.

Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546-4559.

Lomonossoff, G. P., Butler, P. J. G., & Klug, A. (1981) *J. Mol. Biol. 149*, 745-760.

Lown, J. W., Hanstock, C. C., Bleackley, R. C., Imbach, J.-L., Rayner, B., & Vasseur, J. J. (1984) *Nucl. Acids Res. 12*, 2519-2533.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95-117.

McLean, M. J., & Wells, R. D. (1988) *J. Biol. Chem. 263*, 7370-7377.

Nagayama, K., Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321-334.

Naylor, L. H., Lilley, D. M. J., & van de Sande, J. H. (1986) *EMBO J. 5*, 2407-2413.

Naylor, L. H., Yee, H. A., & van de Sande, J. H. (1988) *J. Biomol. Struct. Dynam. 5*, 895-912.

Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson, L. (1987) *Biochemistry 26*, 3718-3733.

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591-616.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455-475.

Noggle, J. H., & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect*, Academic, New York.

Nordheim, A., & Rich, A. (1983) *Nature 303*, 674-679.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187-193.

Patel, D. J., Kozlowski, S. A., Nordheim, A., & Rich, A. (1982) *Proc. Natl. Acad. Sci. 79*, 1413-1417.

Patel, D. J., Shapiro, L., & Hare, D. (1987) *Ann. Rev. Biophys. Biophys. Chem. 16*, 423-454.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498-504.

Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dynam. 4*, 621-649.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327-337.

Scalfi Happ, C., Happ, E., Nilges, M., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 1735-1743.

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914-2916.

Searle, M. S., Hall, J. G., Denny, W. A., & Wakelin, L. P. G. (1988) *Biochemistry 27*, 4340-4349.

Shindo, H., Okhubo, S., Matsumoto, U., Giessner-Prettre, C., & Zon, G. (1988) *J. Biomol. Struct. Dynam. 5*, 913-931.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.

Stevens, W. C., Huang, D. H., Wells, R. D., & Krishna, N. R. (1988) *Nucl. Acids Res. 16*, 7133-7143.

Suck, D., Lahm, A., & Oefner, C. (1988) *Nature 332*, 464-468.

Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854-6865.

Tidor, B., Irikura, K., Brooks, B. R., & Karplus, M. (1983) *J. Biomol. Struct. Dynam. 1*, 231-252.

van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1-38.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature 282*, 680-686.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, Jr., S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765-784.

Chapter IV

The Solution Structure of Phage λ Half-operator DNA using Nuclear Magnetic Resonance, Restrained Molecular Dynamics, and NOE-based Refinement[1]

[1]A version of this chapter has been accepted for publication: Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry* (in press). Reprinted, with permission, from Baleja, J. D., Pon, R. T., & Sykes, B. D. *Biochemistry* (1990, in press). Copyright (1990) American Chemical Society.

A. Introduction

The sequence-specific recognition of DNA by proteins is an important component in the control of cellular processes. The working model of protein-DNA interaction for this thesis is the Cro repressor-O_R3 operator system from bacteriophage λ (for reviews, see Brennan & Matthews, 1989a; Pabo & Sauer, 1984; Takeda et al., 1983; Ptashne et al., 1980) in which the helix-turn-helix DNA binding motif of the protein recognizes a specific sequence of DNA. The Cro repressor is one of a family of structurally related DNA-binding proteins that include the lac repressor (Kaptein et al., 1985), the repressor from phage 434 (Wolberger et al., 1988), and from phage λ (Jordan & Pabo, 1988), and the catabolite gene activator protein (CAP) from Escherichia coli (McKay and Steitz, 1981). These proteins bind as dimers to two-fold symmetric operator DNA sites, with the recognition helix of the DNA binding motif of each monomer fitting into successive major grooves of B-type DNA (Anderson et al., 1981). The DNA is directly read via a network of hydrogen bonds and other contacts between amino acid sidechains and the exposed functional groups in the major groove of DNA. This would suggest that the sequence diversity of DNA need not generate different DNA structures, but merely provide different patterns of hydrogen-bonding in the protein-DNA recognition process.

However, the recent crystallographic determination of the related trp repressor complexed with its operator DNA shows no direct sequence-specific contacts between protein and DNA, suggesting that the trp repressor is recognized indirectly through sequence-specific changes in the geometry of the phosphate backbone to allow the formation of a tight protein-DNA complex (Otwinowski et al., 1988). In addition, although the x-ray structure of the repressor-operator complex from phage 434 shows no direct protein contact at the center of the operator (Aggarwal et al., 1988), substitution of these base-pairs changes the binding affinity of the protein for its DNA (Koudelka et al., 1987; Koudelka et al., 1988). From modelling studies with B-DNA, and from extensive biochemical data, the same situation is likely to exist with the phage λ Cro repressor protein. There are no direct contacts between protein and the center of the O_R3 operator, but substitution of the central base-pair results in slightly reduced affinity of protein for DNA (Ohlendorf et al. 1982, Takeda et al., 1989). These results suggest a sequence-dependent change in conformation of the operator to accomodate for binding to the Cro dimer (Kirpichnikov et al., 1984) or, more generally, that DNA does not merely play a passive role in protein-DNA recognition but instead, by altering its conformation from regular B DNA, is able to pre-form an optimal binding site for its cognate protein, or alternatively, is able to adopt a lower energy conformation in the protein-DNA complex.

In this chapter, the structure of part of the O_R3 operator DNA in the absence of protein is presented as a basis for comparison to the structure of DNA within the protein-DNA complex. We have chosen to study a chemically synthesized ten base-pair DNA fragment using two-dimensional NMR[1] techniques. The sequence comprises the left most ten base-pairs (L10) of the 21 base-pair operator, and also contains the consensus Cro and λ repressor binding site (Kim et al., 1987). This half-operator binds to Cro repressor at a ratio of 2 to 1, with nearly the same binding strength as the full operator (unpublished results), confirming that the complex is specific, as is observed with the lac repressor headpiece:half-operator DNA system (Scheek et al., 1983). First, non-exchangeable proton resonances are assigned in a sequential manner by a combination of two-dimensional NOE (NOESY) and through-bond correlated (COSY) spectroscopies. Assignments are extended to base imino and hydrogen-bonded amino protons by one-dimensional NOE experiments in H_2O. A set of allowed dihedral angles is derived from a qualitative interpretation of COSY intensities and cross-peak patterns (Chary et al., 1988). From time-dependent NOE cross-peak intensities, another set of approximate interproton distances is obtained (Baleja et al., 1990). These two sets of data are then used as the basis for a structure determination by restrained molecular mechanics calculations (Kaptein et al., 1985; Nilsson et al., 1986). Convergence is achieved (rms deviation of 0.8 Å) starting from two quite different initial structures, namely classical A and B type DNA (rms deviation of 5.4 Å).

The distance between two spins is often estimated by assuming inverse proportionality to the sixth root of the NOE cross-peak intensity. Distances so derived are only approximate since the cross-peak intensity due to direct cross-relaxation between spins i and j is modified by additional cross-relaxation with any spin k, especially if spin k exists such that $r_{ik} < r_{ij}$ or $r_{jk} < r_{ij}$. However, NOE cross-peak intensities may be predicted from the structures produced by dynamics calculations and compared directly to the observed intensities, eliminating approximate distance calculation (Chapter II). The structures resulting from restrained molecular dynamics calculations are refined in an iterative manner so as to minimize the difference between the two sets of NOEs (Baleja et al., 1990). The final structures have NOE R factors of less than 0.19, consistent with the observed experimental data. The structure is of the B type DNA family, and exhibits sequence dependent variations in conformational parameters. The full O_R3 operator is derived from the half-operator structure. Although some subtle differences from classical B-DNA are displayed in the pattern of proposed protein contacts in the Cro repressor binding site, it is basically consistent with the most recent Cro repressor-DNA models (Takeda et al., 1989). A comparison is made between the structure of the DNA free in solution, and when bound to the related λ repressor protein (Jordan & Pabo, 1988).

B. Experimental Procedures

1. Sample Preparation

The two deoxyoligonucleotides, d 5'(TCTATCACCG) and d 5'(CGGTGATAGA), were prepared (by R. T. Pon, University of Calgary) on an Applied Biosystems DNA synthesizer using a large scale synthesis procedure (Lee et al., 1988). Coupling yields for the syntheses averaged 95-97%. After deprotection with thiophenol and ammonium hydroxide, synthesis products were purified by chromatography on NACS-20 resin using a 0.25 to 0.45 M NaCl gradient in 12mM NaOH (Lee et al., 1988). Pooled fractions were neutralized with acetic acid and desalted on a Sephadex G-25 column. Electrophoresis on 24% polyacrylamide/7M urea gels and re-chromatography on NACS-20 confirmed product homogeneity.

An aliquot of each strand (at the University of Alberta) was digested to 5' nucleotides for 1 hour with 10μL of ca. 0.1% snake venom phosphodiesterase in 20mM TRIS, 1mM $MgCl_2$, pH 9 buffer. The concentration of each strand was determined by the absorbance of nucleotide mixture using the following molar absorbancies at 260 nm.: dC, 7900; pdC, 6400; dA, 14700; pdA, 15300; dG, 11800; pdG, 11700; dT, 9600; pdT, 7600. Heating equimolar amounts of strands to 85C in 140μL of a 60mM KCl, 20mM KPO_4, 50μM EDTA, pH 7 buffer, and then allowing the solution to cool to room temperature over several hours, ensured strand annealing to form the duplex:

5' (T_1 p C_2 p T_3 p A_4 p T_5 p C_6 p A_7 p C_8 p C_9 p G_{10}) 3' (Strand 1)
3' (A_{20} p G_{19} p A_{18} p T_{17} p A_{16} p G_{15} p T_{14} p G_{13} p G_{12} p C_{11}) 5' (Strand 2)

Solutions were passed over the Na^+ form of Chelex-100 to remove paramagnetic metal ions before lyophilization. Samples were dissolved and re-lyophilized repeatedly in increasing grades of D_2O, and finally taken up in 0.65 mL of 99.997 % D_2O. Final concentrations were 4 mM duplex, 0.2 M KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 20μM EDTA, pH* 7.5 (direct meter reading). One-dimensional ^1H NMR spectroscopy confirmed product homogeneity (Figure IV-1).

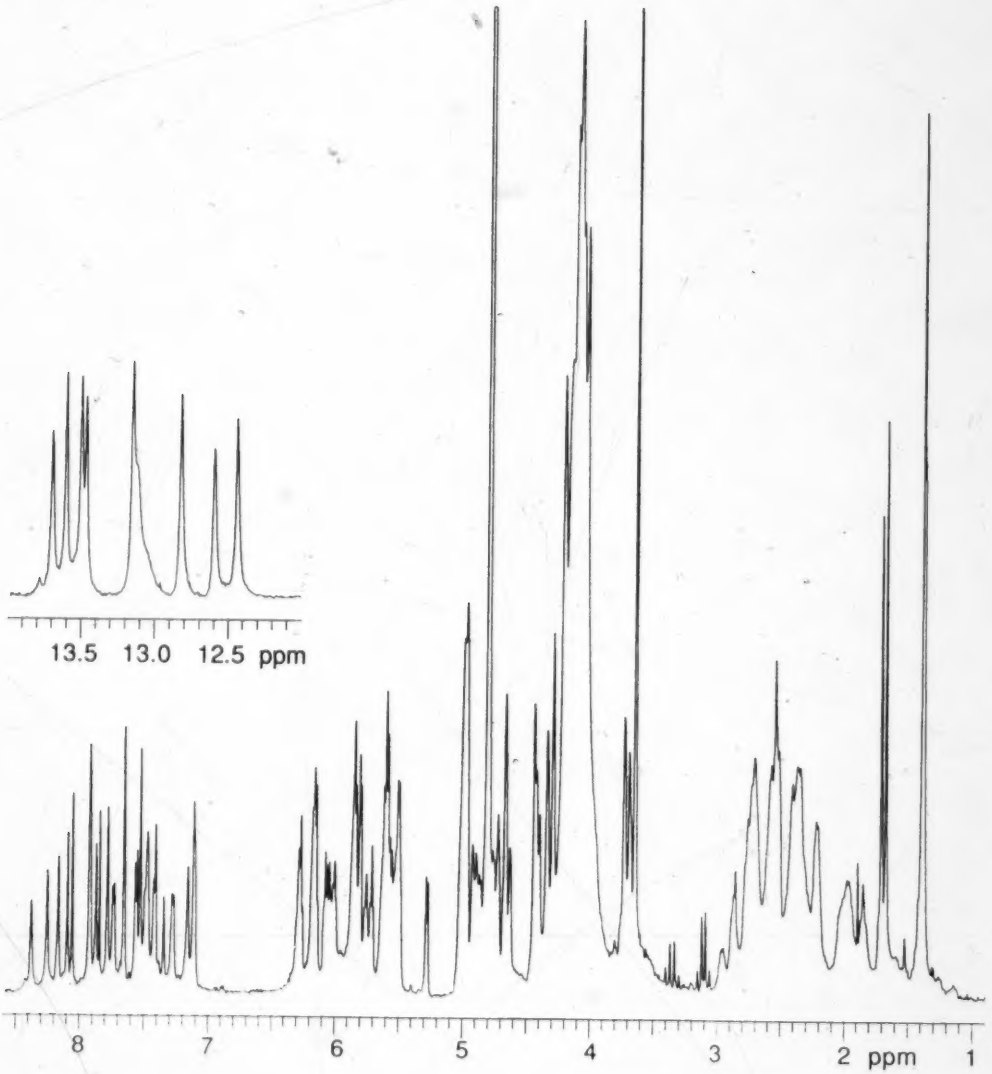


Figure IV-1 One-dimensional ^1H NMR spectrum of d(TCTATCACCG)•d(CGGTGATAGA), L10, in D_2O at 20°C. The inset is the imino proton region spectrum in 85% H_2O, 15% D_2O, 7.7°C.

2. NMR Spectroscopy

All NMR spectra were obtained on a Varian VXR500 NMR spectrometer with an operating frequency of 500 MHz for protons. Phase sensitive two-dimensional spectra in D_2O were taken with 2K complex points with appropriate phase cycling for quadrature detection and for eliminating axial peaks. Streaking along t_1 was reduced by multiplying the first domain time point by a factor optimized near 0.5 (Otting et al., 1986). The residual HOD signal was suppressed by pre-saturation. Spectra were normally taken at 20°C, although spectra at 30°C were useful to confirm assignments at the lower temperature. In particular, double quantum filtered COSY spectra at 30°C had better signal-to-noise ratios, and small chemical shift changes, but were otherwise identical to 20° spectra. Free induction decays were weighted in each dimension to effect slight resolution enhancement and to avoid truncation effects.

NOESY spectra were collected using the hypercomplex method (States et al., 1982). Average mixing times of 50, 100, 150, and 250 milliseconds were used with a random delay of ±10 milliseconds incorporated to suppress zero quantum coherence. 256 t_1 increments were taken and for each t_1 value, 64 scans with an acquisition time of 0.215 seconds were signal averaged. The delay time between scans was 2.1 seconds. Spectra were zero-filled in each dimension so that final transformed spectra were 1K by 1K real data points. NOE intensities were quantified by determining the volume integral of each cross-peak. Nominally empty areas perpendicularly adjacent to each cross-peak were examined for baseline correction.

Through-bond correlations were observed with DQF-COSY spectra (Piantini et al., 1982). For each t_1 value, 64 scans with an acquisition time of 0.26 seconds were signal averaged with a delay time of 2.0 seconds between scans. 512 t_1 increments were taken and final spectra were 2K by 1K data points.

One-dimensional spectra were also taken in 85% H_2O/15% D_2O at 10°C with 1-1 binomial suppression of the H_20 peak (Hore, 1983). For difference NOE spectra (Gronenborn & Clore, 1985), the decoupler power was chosen (γB_2 = 8Hz) to completely saturate the peak during the mixing period while still retaining specificity. Mixing times between 100 and 500 milliseconds were used. 2048 transients with a preceeding delay time of 2.5 seconds and an acquisition time of 0.5 seconds were summed into 13K data points.

3. Restrained Molecular Dynamics

Energy minimization and molecular dynamics (MD) calculations were carried out with the GROMOS program (van Gunsteren et al., 1985; de Vlieg et al., 1986) and force field, which consisted of the usual terms for bonds, bond-angles, sinusoidal dihedral torsion, non-bonded interactions (van der Waals and electrostatics), and harmonic terms to maintain proper planar or tetrahedral geometries, and to which two extra terms representing distance and dihedral restraints were added (Chapter III). The experimental distance and dihedral data sets were based on observations from the two-dimensional NMR spectra (see below).

Starting models were first subjected to 200 steps of steepest descents energy minimization. During the first 10 picoseconds of each MD simulation, values of the distance restraint force constant were increased from 500 to 10000 kJ mol^{-1} nm^{-2} and the dihedral restraint force constant from 5 to 50 kJ mol^{-1} rad^{-2}, as outlined in Chapter III. Molecular dynamics runs with the highest values of CDIS and CDLR were continued to 20 picoseconds in total, and coordinates were averaged over the last 5 picoseconds. Averaged molecular dynamics structures were then subjected to 200 steps of energy minimization to correct distortions in the structure caused by the averaging procedure.

A total of four molecular dynamics runs were performed : (1), with a starting model of L10 in a classical A-type DNA conformation (Arnott & Hukins, 1972), a set of distance restraints, and a set of dihedral restraints; (2), with a starting model in an average B-DNA configuration, and the experimental restraints; (3), with a starting A-DNA model and the sets of distance and dihedral angle restraints that correspond to the same atoms for the experimental restraints, but with values taken from the B-DNA model; and (4), with a starting B-DNA model with A-DNA restraints.

4. Structure refinement

Molecular dynamics calculations result in structures that satisfy the two sets of experimental restraints—the approximate distance set and the dihedral angle set. Distance restraints based on the NOE intensities are most often inaccurate because of spin diffusion effects. In addition, the Watson-Crick base-pairing distance restraints prevent any substantial deviations for the two bases of a base-pair from lying in a single plane, and may affect certain conformational parameters, such as propeller twist. Structures were therefore subjected to further refinement which was directly based on the NOE intensities (without distance restraints) by replacing the effective potential for distance restraints by an E_{NOE} potential for NOE restraints, as was done in Chapters II and III.

The pseudo-energy NOE forces were calculated for NOE intensities at 150 and 250 milliseconds, which was a compromise between less signal-to-noise at shorter mixing times, and greater spin diffusion effects at long times. A total of 294 NOE intensities at 150 milliseconds, and 309 NOE intensities at 250 milliseconds were used for structure refinement. NOE intensities between non-exchangeable protons were calculated assuming homonuclear dipolar relaxation for a macromolecule tumbling isotropically in solution (Chapter II).

To take into account differential motion during refinement, empirical correlation time reduction factors were associated with each proton of the proton pair with values of 0.65, 0.85 and 0.9 used to reflect the increased motion of all thymine methyl groups, sugar 2' and 2" methylene protons, and the 5' and 3' terminal residues, respectively, with an overall correlation time of 3.8 nanoseconds (Chapter II). Structure refinement began with energy minimization using a dihedral force constant of 50 kJ mol^{-1} rad^{-2} and an NOE force constant of 1000 kJ mol^{-1} $(\Delta NOE)^{-2}$ and continued until the total potential energy changed by less than 0.001 kJ per step (about 50 steps). A further 100 steps of minimization with CNOE set to 2000 kJ mol^{-1} $(\Delta NOE)^{-2}$ completed the refinement procedure. Helical parameters of the final structures were analyzed with the programs AHELIX, BROLL, and CYLIN (Fratini et al., 1982; Dickerson et al., 1985).

E. Results

1. Resonance Assignment

A prerequisite for the determination of a solution structure by NOE measurements is the assignment of a resonance to a specific proton of the macromolecule. Resonance lines in nucleic acids can be identified by a combination of COSY and NOESY experiments, which measure respectively, coherent and incoherent transfer of magnetization between individual nuclei. A qualitative interpretation of the NOESY spectra reveals that the L10 decamer is right-handed (Cohen, 1987) and resonances were therefore assigned following the appropriate procedures developed for right-handed helices (Chapter III). Regions of the NOESY and DQF-COSY spectra used for assignment are shown in Figures IV-2 to IV-6. Exchangeable protons were assigned by one-dimensional difference NOE spectroscopy (Gronenborn & Clore, 1985), as shown in Figure IV-7. Resonance assignments are given in Table IV-1.

2. Distance determination

As in Chapter III, distances between protons are obtained using a distance extrapolation procedure. Distances are derived at each mixing time, plotted against the mixing time, and are extrapolated back to zero mixing time as a first order correction for spin diffusion effects:

$$r_{ij} = \lim_{\tau_m \to 0} r_{ij}(\tau_m) = \lim_{\tau_m \to 0} \left\{ \frac{NOE_{ref}(\tau_m) * r_{ref}(\tau_m)^6}{NOE_{ij}(\tau_m)} \right\}^{\frac{1}{6}} \qquad (1)$$

For NOEs involving both base and 1' sugar protons, the cytosine H5-H6 NOE and reference distance of 2.46 Å is used. Remaining distances between non-exchangeable protons are determined using the average $NOE_{ref}*r_{ref}^6$ products for the C2' methylene 2'<->2" proton pair (1.76 Å) and the 1'<->2" pair (2.3±0.1 Å for all sugar puckers). Upper and lower bounds on these distances are estimated from the distance extrapolation curve. For all distances greater than 3.5 Å and involving both an aromatic base and either of the C2' methylene protons, upper bounds are increased by 0.2 Å to account for systematic distance under-determination for this arrangement of protons (Baleja et al., 1990).

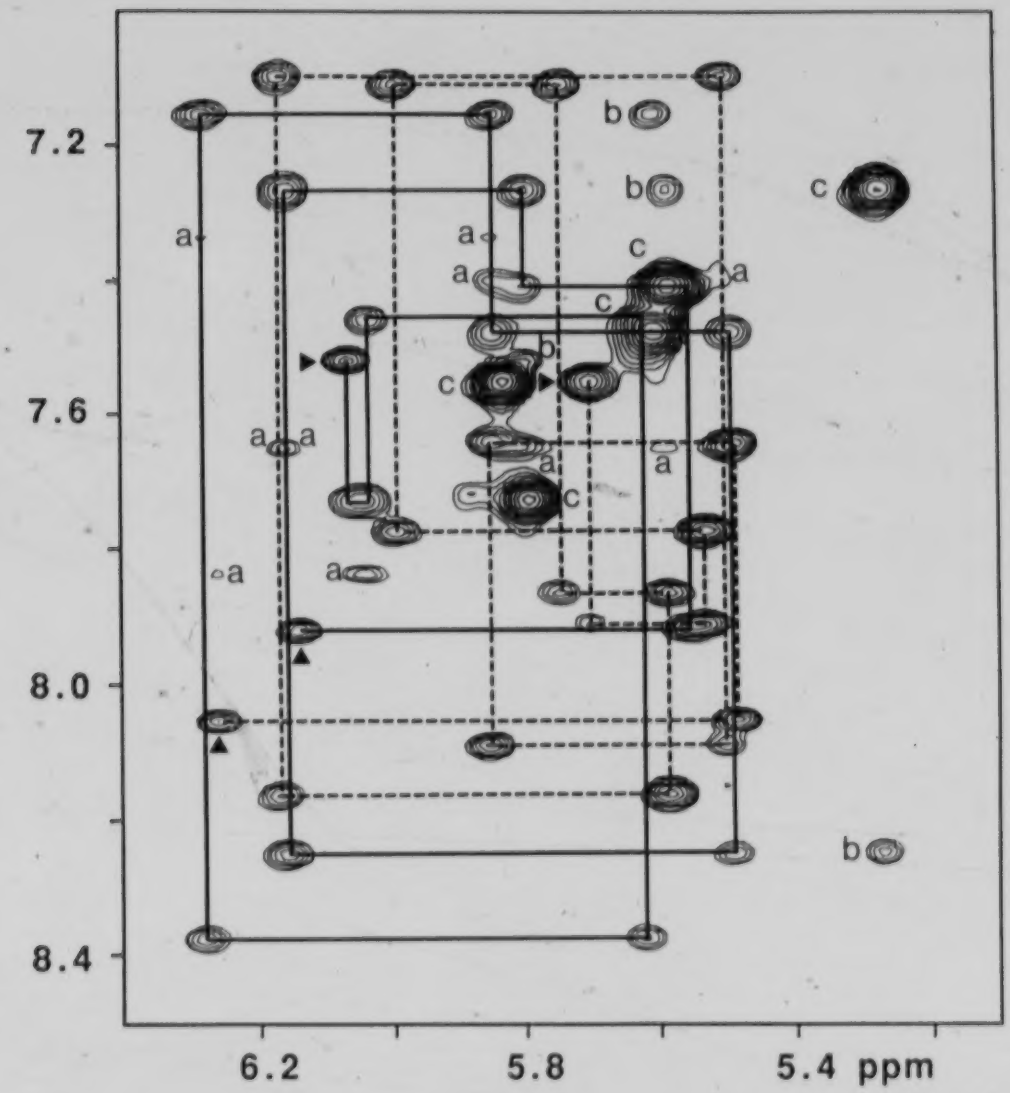


Figure IV-2 Assignment of base and 1' protons of L10 DNA

The mixing time of the 500-MHz NOESY experiment was 250 milliseconds. Sequential connectivites between H6/H8 and 1' protons are shown by the solid line for the Strand 1 (residues 1-10) and by the dashed line for Strand 2 (residues 11-20). (a) Cross-peaks to adenine H2 protons. (b) NOE cross-peaks between the H5 base proton of cytosine and the H6/H8 proton of the preceeding base. (c) H5-H6 cytosine cross-peaks.

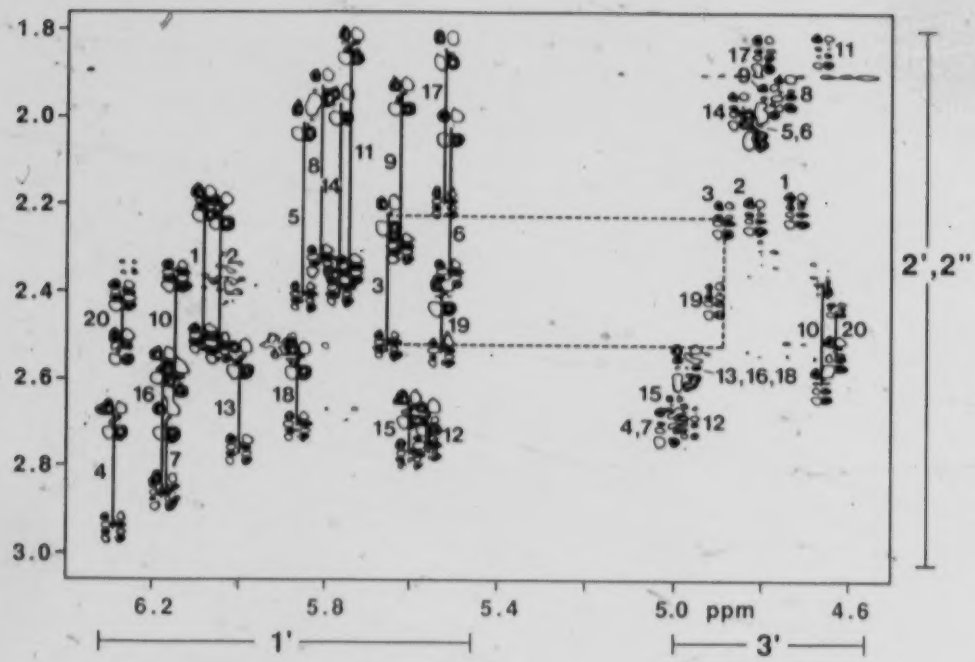


Figure IV-3 Assignment of 2', 2", and 3' protons

The DQF-COSY ^{1}H NMR spectrum was taken at 30°C. Cross-peaks between 1' <-> 2' and 1' <-> 2" protons are indicated by residue number. Except for the 3' terminal, 2' protons resonate upfield to the 2" proton of the same residue. The dashed line illustrates the extension of the assignment from 2',2" protons to the 3' proton for residue T_3. As observed with other residues, the 2" <-> 3' cross-peak is absent. 2' <-> 3' cross-peaks are noted by residue number.

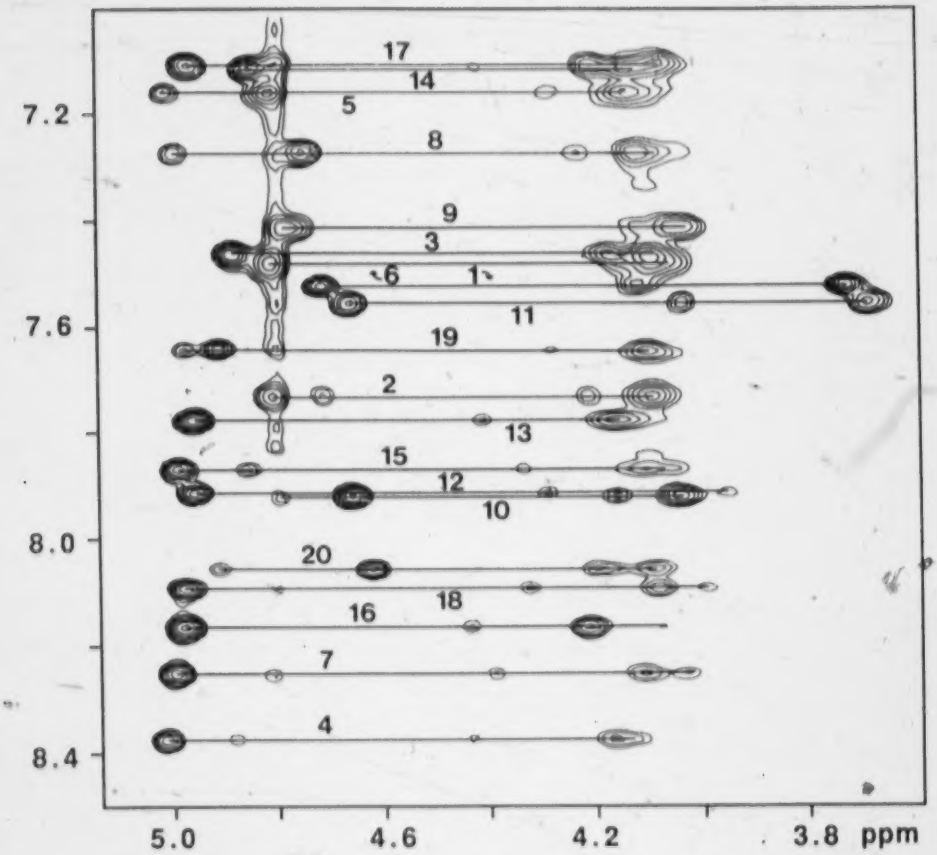


Figure IV-4 Assignment of 3', 4', 5' and 5" protons

Base<->3', 4', 5' and 5" region of a 250 millisecond mixing time NOESY spectrum at 20°C. H6/H8 proton resonance positions are indicated by residue number. Cross-peaks represent intra-residue correlations, except for the weaker base<->3' cross-peaks which are NOE connectivities from the base proton to the 3' proton of the previous (5') residue. No cross-peaks were observed to adenine H2 protons.

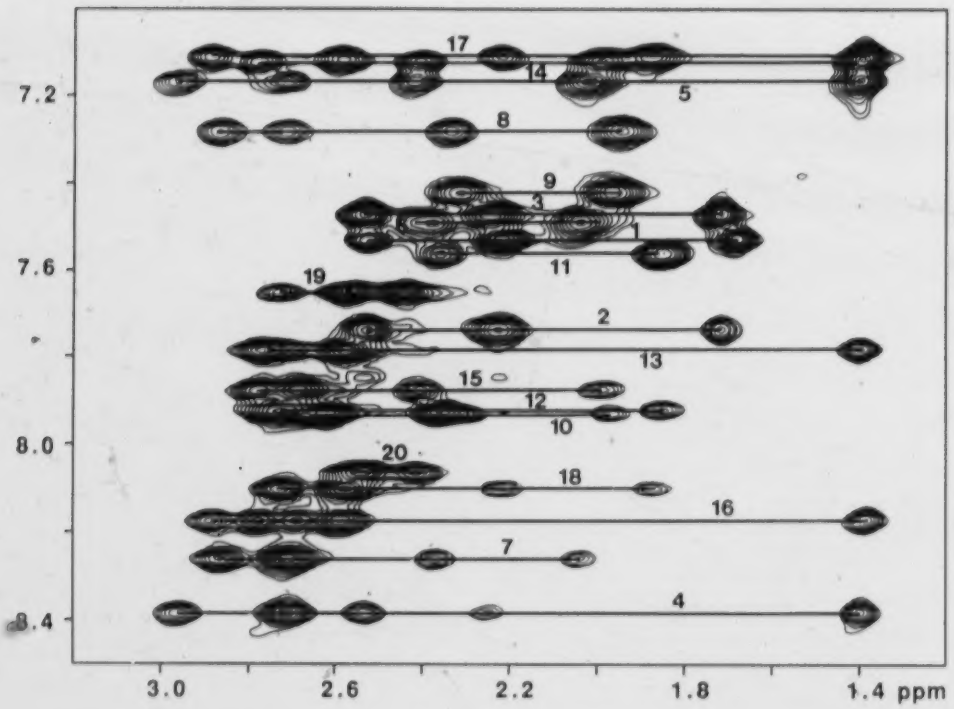


Figure IV-5 Base<->2',2" region of the NOESY spectrum

Base proton resonance positions are indicated by residue number. NOESY walks down each strand, can be followed in a manner similar to that indicated in Figure IV-2, confirming assignments.

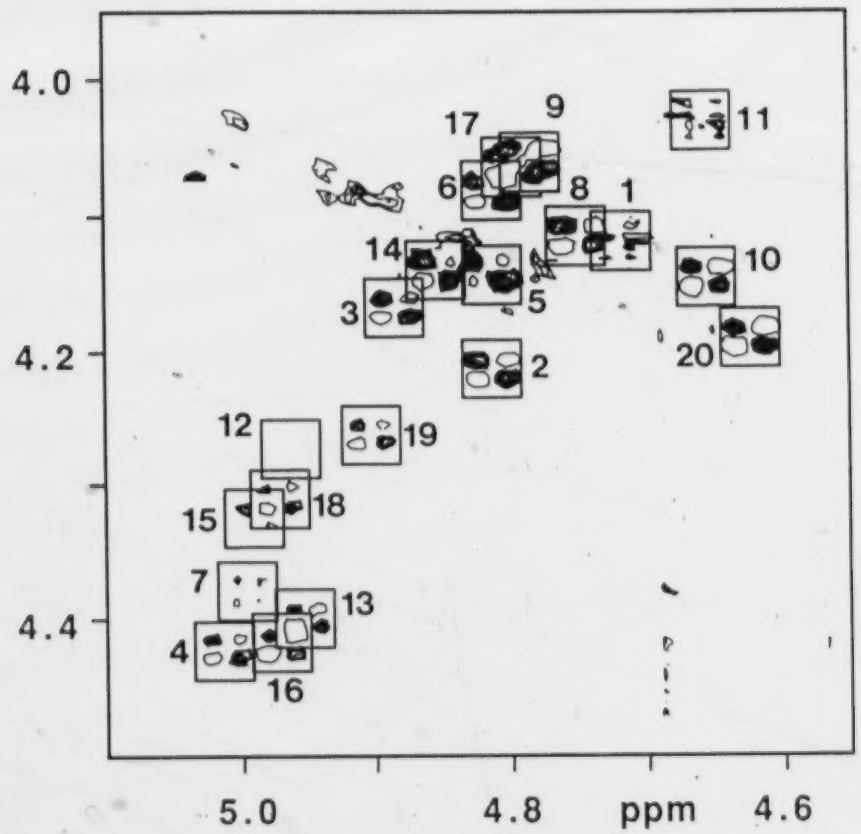


Figure IV-6 3'<->4' region of the DQF-COSY spectrum at 30°C

The single missing correlation is G_{12} 3'-4'. This is most likely due to the 3' being coupled into a second order 2',2" methylene proton spin system ($|J| / \Delta\delta \cong 14Hz / (0.05ppm * 500MHz) \cong 0.56$), causing a more complicated cross-peak pattern with an inherent lower signal intensity. This cross-peak can be visible just above the noise level of the spectrum if the vertical gain of the spectrum is increased (which also results in the residual HOD t_2 streak at 4.68 ppm obscurring most of the other peaks).

Figure IV-7 One-dimensional Nuclear Overhauser Effect experiments in 85% H_2O, pH 6, 7.7°C. The mixing time was 500 milliseconds. Irradiation at: (A) 13.69 ppm, T_3 H3; (B) 13.60 ppm T_{14} H3; (C) 13.50 ppm, T_5 H3; (D) 13.47 ppm, T_{17} H3; (E) 13.15 ppm, G_{12} H1; (F) 12.82 ppm, G_{13} H1; (G) 12.60 ppm, G_{19} H1; (H) 12.44 ppm, G_{15} H1. Imino protons resonate between 12 and 14 ppm, hydrogen-bonded cytosine N4 amino proton between 8 and 8.6 ppm, and adenine H2 protons between 7 and 8 ppm. NOE intensities are labelled with the residue number, and resonance assignments are given in Table IV-1.

Table IV-1 Proton Chemical Shift Assignments for L10[a]

Residue	H6/H8	H2/H5/ CH_3	1'	2'	2"	3'	4'	5', 5"[b]	NH[e]
T_1	7.53	1.67	6.08	2.20	2.52	4.72	4.14	3.73 , 3.73	[f]
C_2	7.74	5.80	6.04	2.22	2.52	4.83	4.23	4.10 , 4.08[c]	8.46
T_3	7.47	1.71	5.63	2.23	2.53	4.89	4.18	4.11 , 4.14[c]	13.69
A_4	8.38	7.42	6.29	2.70	2.96	5.03	4.45	4.15 , 4.18	-
T_5	7.17	1.39	5.86	2.00	2.41	4.83	4.17	4.15 , 4.30	13.50
C_6	7.49	5.61	5.50	2.03	2.37	4.82	4.11	[d] , [d]	8.41
A_7	8.26	7.68	6.17	2.70	2.86	5.00	4.40	4.04 , 4.12	-
C_8	7.27	5.28	5.81	1.93	2.32	4.76	4.14	4.23 , 4.09	8.11
C_9	7.42	5.60	5.57	1.95	2.29	4.79	4.07	4.04[c], 4.02[c]	8.64
G_{10}	7.93	-	6.15	2.61	2.36	4.67	4.17	4.07[c], 4.01[c]	13.10
C_{11}	7.56	5.84	5.71	1.83	2.35	4.67	4.05	3.69 , 3.69	-
G_{12}	7.92	-	5.54	2.69	2.74	4.97	4.31	4.07 , 3.96	13.15
G_{13}	7.79	-	6.00	2.58	2.77	4.97	4.43	4.16 , 4.18	12.82
T_{14}	7.12	1.39	5.76	1.98	2.39	4.86	4.16	4.22 , 4.10[c]	13.60
G_{15}	7.88	-	5.60	2.67	2.77	5.00	4.35	[d] , [d]	12.44
A_{16}	8.17	7.68	6.18	2.58	2.88	4.99	4.44	4.23 , 4.25[c]	-
T_{17}	7.11	1.37	5.51	1.86	2.20	4.80	4.08	4.22 , 4.15[c]	13.47
A_{18}	8.10	7.35	5.86	2.57	2.72	4.98	4.34	4.00 , 4.08	-
G_{19}	7.67	-	5.50	2.44	2.57	4.92	4.29	4.11[c], [d]	12.59
A_{20}	8.07	7.87	6.27	2.53	2.39	4.63	4.22	4.18 , 4.09	-

[a]Chemical shifts (ppm) are relative to disodium-2,2-dimethyl(-2-silapentane-5-sulphonate). Non-exchangeable proton assignments are at 20°C and exchangeable proton assignments are at 7.7°C. [b]5' and 5" protons are not assigned stereospecifically. [c]Tentative assignment. [d]Unassigned 5' and 5" resonances are between 4.05 and 4.25 ppm. [e]Assignments are for the H1 imino proton of guanine, the H3 imino proton of thymine, or the hydrogen-bonded amino proton (on N4) of cytosine. [f]No unambiguous resonance was observed for T_1 because of spectral overlap, rapid exchange with H_2O, and poor signal intensity.

NOE intensities are calibrated for exchangeable protons by assuming an average 3.4 Å distance for imino-imino NOEs between sequential residues, 2.5 Å for the Guanine H1 <-> Cytosine H4a (hydrogen-bonded amino proton) interstrand proton pair, and a 3.0 Å distance for the Thymine H3 <-> Adenine H2 interstrand proton pair. Increased upper distance estimates, and decreased lower distance estimates were used for the distances involving exchangeable protons because of measurement at a lower temperature, and loss of magnetization due to exchange with H_2O.

For overlapping peaks, lower bounds for the distances corresponding to the proton pairs involved could be estimated by using all of the observed NOE intensity for each pair. 322 upper distance bounds and 330 lower distance bounds comprised the set of distance restraints determined from the NOE intensities for input into molecular dynamics calculations. NOE measurements and distances are given in the experimental restraints tables of Appendix 5. Little residue to residue variation is seen at short mixing times for interproton vectors of fixed length, in either distances or NOE intensities, indicating only a small amount of increased motion for terminal residues on the nanosecond time-scale. At longer mixing times, however, NOE intensities are dissimilar for terminal residues. Much of this effect can be modelled by calculations taking spin-diffusion into account, indicating that the lower spin density at the ends of the molecule results in less spin diffusion and the NOE intensity for short covalently bound distances is therefore lower.

3. Glycosidic dihedral angles

The geometry of the five-member sugar ring of DNA can be described by five torsion angles ν_0 - ν_4. Because of ring closure, the values of ν_0 - ν_4 are interrelated:

$$\nu_n = \nu_{max} * \cos [P + 144* (n-2)] \qquad n=0,4 \qquad (2)$$

where ν_{max} is the maximum amplitude of the sugar ring pucker, and P is the sugar pseudo-rotational angle (Altona & Sundaralingam, 1972). The magnitude of three-bond coupling constants between protons is dependent on the intervening dihedral angle, and therefore reflects the pseudorotational angle that specifies the conformation of the sugar ring (Hosur et al., 1986).

Excepting 3' termini, 2"<->3' correlations are absent in the COSY spectrum (Figure IV-3). This indicates that the sugar pseudorotational angle is between 100 and 250 degrees for all non-(3')terminal sugar rings (Hosur et al., 1986). Furthermore, the separation of the outer lines in cross-peaks involving the 1' proton is greater than 14 Hz in the 1' frequency

dimension. This indicates the sum of coupling constants to the 1' proton for all residues to be greater than 14 Hz, which further limits the pseudorotation angle to lie between 100 and 200 degrees (Chary et al., 1988; Rinkel & Altona, 1987). The observation of 3'<->4' correlations (Figure VI-6) for all residues narrows the pseudorotation angle to be between 105 and 175°. Allowed ranges of glycosidic dihedral angles are obtained for each nucleotide (Appendix 5) from the coupling constant data assuming a v_{max} of 35° (Rinkel & Altona, 1987), and by using equation 2.

4. Right-handed DNA helix restraints

To preserve the right-handed character of the DNA during the molecular dynamics calculations, it was sometimes necessary to constrain backbone dihedral angles to be in a broad allowed region of torsional angle space (Gronenborn & Clore, 1989). The allowed angles (α, -90 to -30°; β, < -145° and >135°; γ, 30 to 90°; ε, < -60° and >140°; ζ, < -45° and > 150°) are derived from a table of conformation angles found in the different DNA types (Suzuki et al., 1986) and from considering individual variations in right-handed helices from X-ray crystallographic studies (Dickerson et al., 1985). These right-handed helix restraints would cause no violations for any of the average A, B, alternating B, C, D, or wrinkled D DNA forms, nor for any individual angles found in the best studied single crystal X-ray structures of B-DNA (Dickerson & Drew, 1981; Privé et al., 1987), except for two β angles (of residues G_3 and C_{12}) in the B-DNA form of a phosphothioate analogue of DNA (Cruse et al., 1986).

Base-pairs were kept Watson-Crick hydrogen-bonded by distance restraints between bases. These were as listed in Chapter III and gave 120 distance restraints in addition to the 652 noted above.

After approximately ten picoseconds of molecular dynamics, the major groove had a tendency to collapse on two occasions - when starting from B type DNA with experimental restraints, and when starting from A type DNA with B-DNA restraints. A structure formed with a lower van der Waals energy (by 100 kJ), but with also an increased total distance restraint energy (by 30 kJ). The distance restraint force constants could be increased in order to dominate the van der Waals contribution, although an increased force on inaccurate distance restraints would likely cause distortions in the structure. Alternatively, further experimentation could have been carried out with the electrostatic and van der Waals interactions, for example, by adding counterions to the phosphates. Instead, a few 'generous' distance restraints were added between C1' atoms on opposite sides of the major groove (Appendix 5). For example, the distance between T_5 C1' and G_{19} C1' is

12.4 and 9.7 Å in A and in B DNA respectively. During the MD run, these atoms are required to be greater than 8 Å apart. No violations occur with either the right-handed DNA dihedral angle set nor with the major groove distance restraints in any of the structures presented in this paper.

5. Molecular dynamics calculations

Several small alterations to the GROMOS force field were made in order to be more consistent with the nucleic acid force field of the CHARMM molecular mechanics program (Nilsson & Karplus, 1986). The normal van der Waals radius on united methylene carbons was reduced from 2.22 to 2.10 Å to avoid steric clashes between C2' and one of the oxygens on the 3' phosphate. Methine carbons were given van der Waals radii of 2.05 Å. Corresponding 1-4 van der Waals interactions were left unaltered. The effect of solvent was approximated by a $1/\varepsilon r$ screening function where r was the separation of the charged groups in Å (Brooks et al., 1983) and ε was equal to 4 (Weiner et al., 1984; Cuniasse et al., 1989). The net charge on each phosphate group was reduced to -0.32e (Tidor et al., 1983; Nilsson et al., 1986).

The molecular dynamics protocol and force field are tested following the method of Nilsson (Nilsson et al., 1986). A set of distance restraints and dihedral angles (between the same atoms for which there are corresponding experimental measurements and with the same error estimates) are taken from classical B-DNA and are applied to a MD simulation starting from A-DNA and *vice versa*. In both cases, the expected transition occurs with the A to B structure (ABI) having an rms deviation of 1.8 Å from B DNA, and the B to A structure (BAI) has an rms deviation from A DNA of 1.5 Å (Table IV-2). Note that although these distances are taken from idealized structures, they also are from canonically straight DNA with no adjustments made for sequence dependent variations.

The two starting structures (A and B-type DNA models) for the MD runs with experimental restraints are depicted in the top half of Figure IV-8. Application of molecular dynamics produces structures MD-A and MD-B. They have an rms atomic deviation of less than 0.8 Å, which is comparable to the fluctuations during the averaging over the last 5 picoseconds of the MD run. This indicates that molecular dynamics has sampled conformational space well, and that both MD runs have converged essentially to the same structure (Scalfi Happ et al., 1988). Both structures agree with the experimental dihedral and distance restraints equally well, as indicated by the pseudo-potential energies listed in Table IV-3.

Table IV-2 Atomic rms differences (Å) between L10 structures[a]

	A-DNA	B-DNA	R-A	R-B	BAI	ABI
A-DNA		5.43	4.82	4.14	1.49	6.31
B-DNA	4.69		1.30	1.82	5.01	1.82
R-A	4.28	1.08		0.80	4.39	2.29
R-B	3.66	1.55	0.74		3.77	2.88
BAI	1.30	4.49	4.06	3.48		5.66
ABI	5.54	1.38	1.88	2.44	5.17	

[a]A-DNA and B-DNA are starting structures with regular A and B geometries, respectively. R-A and R-B are refined structures after distance restrained molecular dynamics and NOE-based refinement. BAI is the MD structure resulting from applying A-DNA distance restraints to a B-DNA model. ABI is the MD structure resulting from applying B-DNA distance restraints to an A-DNA model. Numbers above the diagonal indicate the rms atomic deviation between two structures. Numbers below the diagonal show the comparison for the middle eight base-pairs.

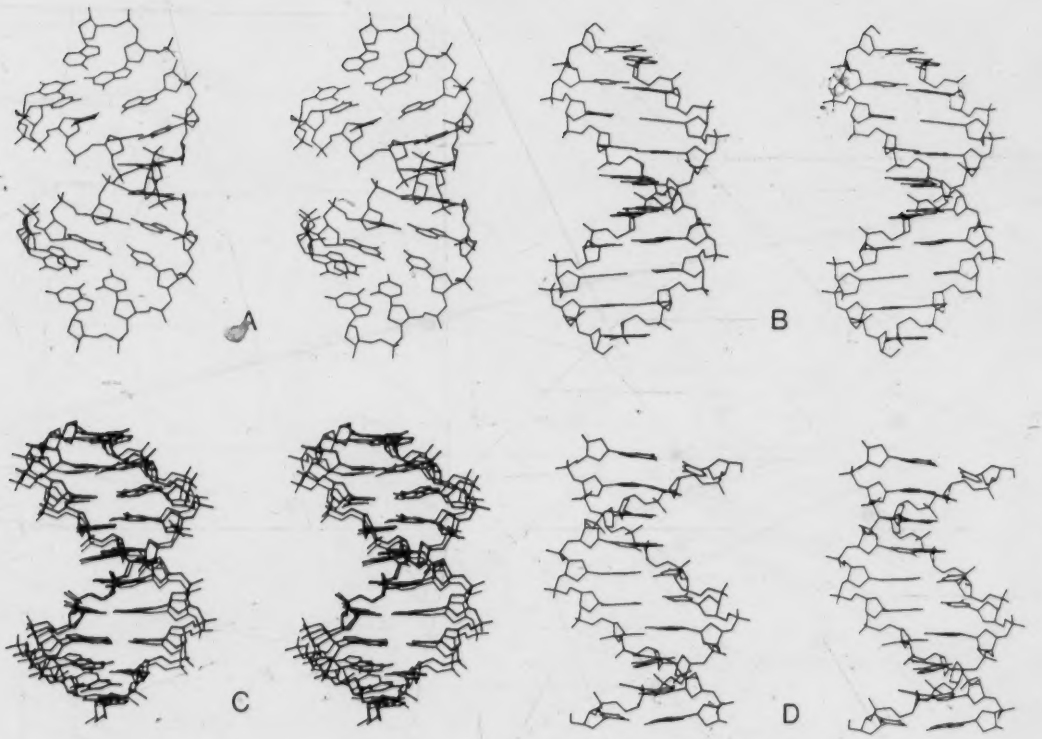


Figure IV-8 Structural models for L10 DNA

(A) Initial model structure in a classical A-DNA conformation. (B) Initial model structure in a B-DNA geometry. (C) Two superimposed final refined structures, R-A and R-B, resulting from restrained molecular dynamics and NOE-based refinement. R-A is shown in bold. (D) The average of structures R-A and R-B. Structure R_{ave} has been rotated by 90° to provide a view into the major groove. Protons have been omitted to preserve clarity in the diagrams.

Table IV-3 Potential energies (kJ/mol) for L10 structures[a]

	Total	elec.	vdW	bond	angle	imp. dihedral	torsion	distance restraint	dihedral restraint	NOE restraint
A-DNA$_m$	-858	-325	-1855	26	220	40	705	240	92	-
B-DNA$_m$	-981	-321	-1787	28	243	18	800	37	0	-
MD-A	-1155	-328	-1934	24	209	45	752	73	4	-
MD-B	-1161	-330	-1924	26	209	46	740	70	4	-
R-A	-1168	-321	-1985	15	223	136	790	-	1	124
R-B	-1166	-322	-1970	15	229	139	775	-	1	125
R$_{ave}$	-1181	-319	-1995	14	224	132	786	-	1	125

[a]Energies are given for initial A and B-DNA models after energy minimization with distance restraints, with CDIS equal to 500 kJ mol^{-1} nm.$^{-2}$ and CDIH at 5 kJ mol^{-1} rad^{-2}. For the remaining structures, CDIS and CDIH are 10000 kJ mol^{-1} nm.$^{-2}$ and 50 kJ mol^{-1} rad^{-2}. MD structures -A and -B result from energy minimization of the average structures from the MD runs starting with the initial A and B- DNA models. For refined (R) structures, the tabulations for total energy exclude the contributions from non-exchangeable protons.

6. Structure refinement and assessment

Each of these structures is then refined separately, replacing the distance restraints with NOE restraints (Baleja et al., 1990). The refined structures, R-A and R-B, again have similar energies, and have an atomic deviation of less than 0.8 Å. These two refined structures are fit to each other (over all atoms) in Figure IV-8C. The bases superimpose best, having an rms deviation of 0.6 Å. The ribose sugars have rms deviations of 0.86 Å, and the phosphate-backbone, 1.05 Å. The less precise determination of the backbone may be due to increased motion (Hogan & Jardetzky, 1980), although there is less data giving direct information on the backbone conformation. Both structures are B-type in character, and have rms deviations approximately 1.5 Å from canonical B-type DNA. Excluding the terminal base-pairs, they exhibit 10.3±0.2 base-pairs per turn of the helix, with an average twist angle of 35.3° and an average spacing between base-pairs of 3.4 Å. The structures bend slightly into the major groove, with the R-B bending somewhat more and being more A-DNA like, despite starting from B-type DNA (Table IV-2). If either half of the molecules are fit to each other (top or bottom 5 base-pairs), the rms deviation for the superimposed atoms is 0.65 Å, and the rms deviation for the remainder is 1.3 Å.

The convergence of the DNA structures from the widely different starting structures indicate that the final refined structures, R-A and R-B, provide reasonable representations of the structure of the half-operator in solution and therefore are discussed in some detail below. The refinement procedure results in little change from the average MD structure (less than 0.1 Å rms deviation). This is consistent with structures that are energetically near a global mininum. Long restrained molecular dynamics simulations, which sample much more conformational space than energy minimization, could not be undertaken because of the computer time required (Baleja et al., 1990).

The NOE residual factor R is used to monitor the fit of the NOEs calculated from a structure to observed NOE intensities (Chapter II):

$$R = \frac{\Sigma | NOE_{obs} - NOE_{calc} |}{\Sigma NOE_{obs}} \qquad (3)$$

where the summation runs over the number of observables (294 NOE intensities at 150 milliseconds and 309 NOEs at 250 milliseconds). To keep an analogy with the standard crystallographic R factor, weighting factors based on the standard deviation in the NOE intensity are not included. Figure IV-9 shows the improvement of the NOE R factor as

molecular structure determination proceeds. The first reduction in the R factor arises from energy minimization of initial models A and B. During molecular dynamics calculations, the R factor steadily decreases for both runs to a value near 0.27. Up to 20 picoseconds, the comparison between observed and calculated NOE intensities is not iterated upon directly, but here the structure is attempting to meet distance restraints inferred from the NOE intensities. Structures obtained by averaging over the last 5 picoseconds of the molecular dynamics calculations produce structures MD-A and MD-B, each with NOE R factors of about 0.24. Final refinement is based on the comparison between observed NOEs and NOE intensities calculated using a full matrix analysis procedure and the differential motion model. After energy minimization the NOE R factor is less than 0.19 for both structures R-A and R-B. Such an R factor (Chapter II) indicates that the structure is consistent with the NOE data within reasonable experimental error.

7. Structural features

As suggested by the overall shape of the molecules and the similar energies and NOE R factors, convergence to the same structure is also seen in details of the spatial relationships of the base-pairs. Four of these conformational parameters, helical twist, roll, propeller twist, and difference in δ torsion angles across the base pair (Figure IV-10), are often predicted from a set of rules based on the crystal structures of the B-DNA dodecamer and other A-type DNA molecules (for a discussion see Dickerson, 1983; Dickerson et al., 1985). Calladine's rules (1982) predict only the roll angle well, which is the rotation that the base-pair makes with respect to its adjacent base-pair about its long axis perpendicular to the overall helix axis. The generally poor correlations for the other parameters has been observed in a number of studies, suggesting that the Calladine steric clash model is insufficient to account for all local helix perturbations (Dickerson et al., 1985; Scalfi Happ et al., 1988; Nilges et al., 1987; Lefèvre et al., 1987; Privé et al., 1987). For example, the twist angle is anti-correlated about base-step 3-4. The interstrand purine-purine clash is avoided at this pyrimidine-purine step instead by base-pair roll, a large negative slide (the distance between midpoints of the base-pairs when viewed down the helix axis) and by increasing the rise between base-pairs slightly (Figure IV-11). The twist angle, propeller twist angles and the difference in δ torsion angles across the base-pair show less dependence on sequence than is observed in crystal structures, where external forces can result in the perturbation of the DNA structure. For example, although the precision of

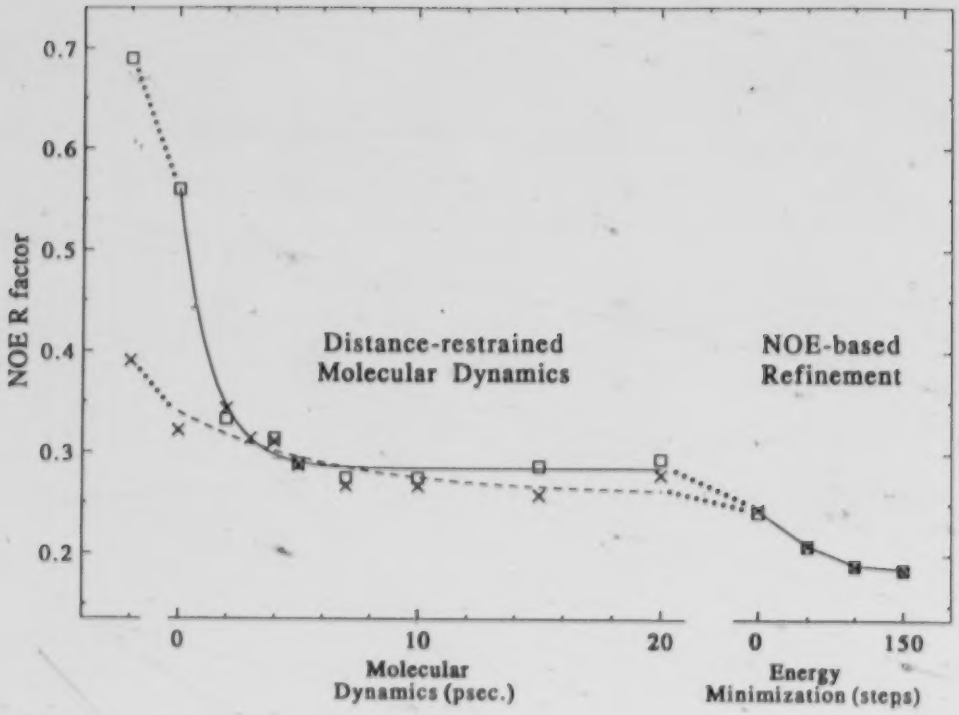


Figure IV-9 Improvement in the NOE R factor as structure determination proceeds
Initial A-DNA models (□) and B-DNA models (x) are first subjected to energy minimization with approximate distance restraints. Twenty picoseconds of distance restrained molecular dynamics reduces the R factor to near 0.27. Averaging over the last 5 picoseconds (second dotted line) results in structures with R factors of 0.24. Structures are refined by comparing observed NOE intensities to NOEs calculated using a full matrix analysis procedure, and minimizing the difference between the two sets of NOEs.

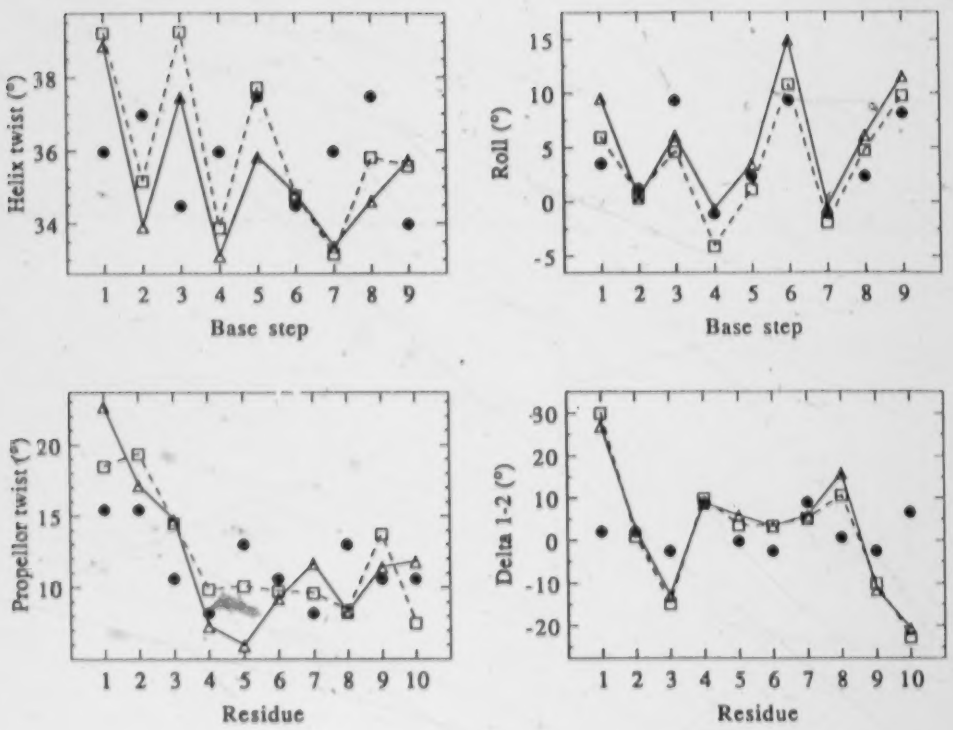


Figure IV-10 Variation in base roll, global helical twist, propellor twist, and $\delta_{1\text{-}2}$, the difference of the C_4' - C_3' torsion angle at each end of a base-pair, for the refined structures R-A (-- □ --) and R-B (—Δ—). Solid circles indicate the predictions based on Calladine's (1982) rules. Precise descriptions for the conformational parameters are given by Dickerson et al. (1985).

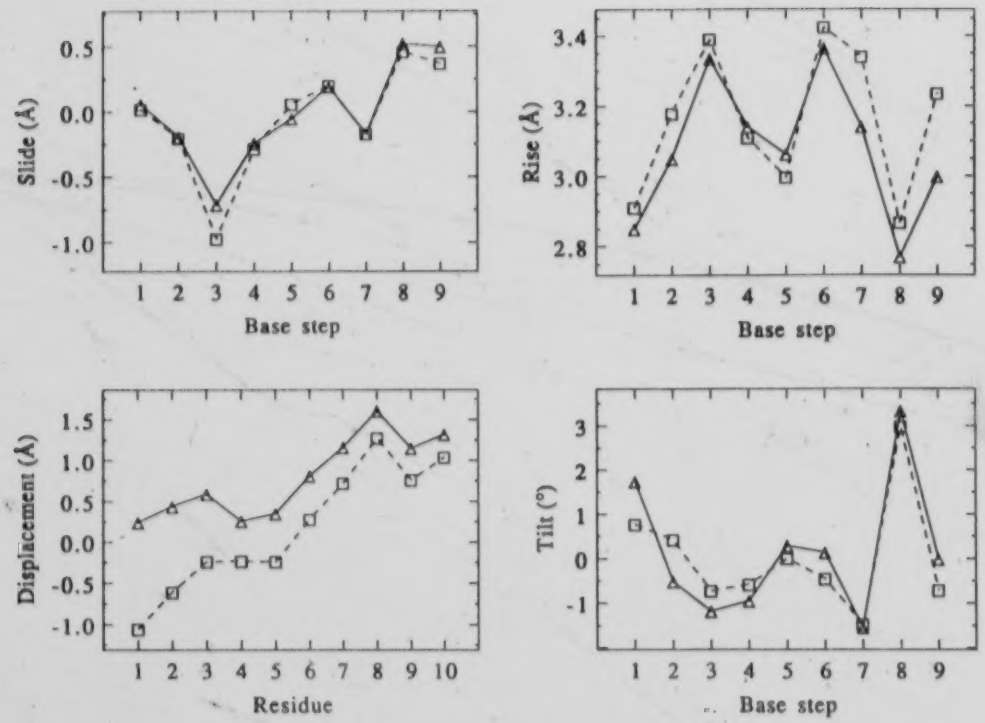


Figure IV-11 Variation in rise, displacement, slide and tilt angles for the refined structures R-A (-- □ --) and R-B (—Δ—).

crystallographic determinations cannot be disputed when compared to other methods, DNA oligomers of mixed sequence have a marked tendency to crystallize in a A-type DNA form, whereas the B-type DNA is clearly preferred in solution (Dickerson et al., 1985).

The distance from the C6-C8 vector of the base-pair to the helix axis, the displacement, averages +4 Å for A-DNA and -0.2 Å for B-DNA. Although certainly not large in magnitude here (Figure IV-11), slightly more positive values for R-B reflect part of its A-DNA type character. The small displacements and tilt angles, (a tilt angle being the angle that the normals of the adjacent mean base-pair planes make with each other), indicate that local twist axes are nearly the same, and are coincident with the global twist axis. This is less true for the terminal base-pairs. A reduction in the network of experimental restraints at the ends of the duplex allows the structure to be more at the mercy of any inadequacies in the molecular force field description. More importantly, chain termination effects causes a discontinuity in the electrostatic and van der Waals interactions of the peripheral base-pairs, likely causing some structural distortions (Grütter et al., 1988).

It is clear that R-A and R-B are essentially equivalent structures. Therefore these two structures are averaged and energy minimized to form a structure R_{ave}, which has, of course, the same R factor as R-A and R-B (0.185). This structure is shown in Figure IV-8D, and has been rotated by 90° from that of Figure IV-8C to present the view from a different perspective. Torsion angles for this structure are shown in Table IV-4. Angles for individual structures, R-A and R-B, varied by less than 3° from the values listed here, except for the 5' terminal γ angles (±5 °). The angles are near the averages observed in the crystal structures of B-DNA sequences (Dickerson et al., 1985; Cruse et al., 1986; Dickerson & Drew, 1981; Privé et al., 1987). Less diversity is displayed for each type of dihedral angle, which may be due to averaging in solution. Note that the initial models chosen are derived from fibre diffraction data, and although adequate as initial models, do not necessarily have realistic dihedral angles. The fact that both determinations of L10 in solution by NMR arrive at a reasonable conformation give credence to the methodology employed.

However, the ε and ζ angles appear to be slightly different from the average angles from single crystal structures. This is, in turn, an average between the B_I trans-gauche$^-$ conformation (for the ε and ζ torsion angles, respectively) and the less common B_{II} gauche$^-$ -trans conformation seen in the crystallographic determinations. The average B_I ε angle observed in crystal structures of B-DNA is near -173 degrees, and the ζ angle is about -85 degrees, which are the angles observed for this decamer in solution. It is not surprising that the B_I conformation should be the one exclusively preferred in solution, as the B_{II} conformation tends to lock the preceeding sugar ring in a C2' endo conformation and

Table IV-4 Dihedral Angles for L10

Residue	χ	α	β	γ	δ	ε	ζ	P
T_1	-126.6			62.1	136.9	-174.5	- 86.7	155.3
C_2	-107.4	-72.4	187.1	55.8	124.1	-180.9	- 88.9	135.1
T_3	-115.0	-65.7	174.9	57.6	112.4	-177.6	- 96.8	117.7
A_4	-105.9	-65.9	172.1	62.5	120.9	-185.5	- 93.1	129.1
T_5	-115.9	-62.3	172.1	61.5	122.1	-177.8	- 98.7	129.7
C_6	-111.1	-67.3	175.1	61.5	121.5	-175.0	- 91.8	126.0
A_7	-111.2	-72.4	174.7	57.7	117.1	-181.5	- 99.3	159.5
C_8	-109.6	-68.5	177.4	60.5	131.3	-184.3	- 89.5	143.8
C_9	-116.0	-63.6	176.6	59.4	116.4	-172.2	- 90.5	122.0
G_{10}	-116.8	-80.7	179.3	50.6	103.2			96.9
C_{11}	-115.6			67.4	124.9	-173.5	- 91.0	134.0
G_{12}	-106.5	-74.7	182.7	55.9	127.1	-178.8	-102.3	138.1
G_{13}	-116.4	-71.5	182.8	55.6	118.1	-178.9	- 84.0	125.6
T_{14}	-113.1	-63.9	170.7	57.1	111.7	-174.7	- 91.5	116.1
G_{15}	-108.3	-67.0	174.9	55.7	117.0	-185.4	- 92.1	125.7
A_{16}	-112.5	-64.0	178.2	56.7	118.8	-176.9	- 92.1	125.8
T_{17}	-116.9	-62.7	169.7	59.1	111.5	-173.4	-104.1	116.4
A_{18}	-102.5	-61.1	166.3	64.5	126.1	-177.6	- 91.7	136.1
G_{19}	-106.7	-67.2	171.1	58.7	122.3	-178.1	-102.7	130.0
A_{20}	-100.8	-71.0	166.7	61.2	108.3			134.7
Ave.	-113.0	-64.9	175.1	59.0	119.6	-178.1	-89.5	128.2
B-DNA	- 98.0	-46.0	-147.0	36.0	156.0	-195.0	-96.0	192.0
A-DNA	-154.0	-87.0	-152.0	45.0	83.0	-182.0	-47.0	14.0
Dodec.[a]	-117.0	-63.0	171.0	54.0	123.0	-169.0	-108.0	-

[a]Average dihedral angles for the DNA dodecamer, CGCGAATTCGCG, (Dickerson & Drew, 1981). Numerical values were not reported for the sugar pucker which ranged from C2'-endo (P=162±18°),through C1'-exo (P=126±18°) to O1'-endo (P=90±18°), with one residue, G24, C3'-endo (P=18±18°).

restricts mobility (Privé et al., 1987; Dickerson & Drew, 1981). The B_{II} conformation appears to be an artifact of crystal packing forces (Dickerson et al., 1987). Note also that β angles following a B_I phosphate conformation average 176° in B-DNA crystal structures, which is the average observed here in solution.

Sugar rings are centered on a C1' *exo* range of pseudo-rotational angle, except the 3' terminal residues, 10 and 20, which are O1' *endo* (Altona & Sundaralingam, 1972). No attempt was made to interpret our results for conformational averaging between C2' *endo* and C3' *endo*, or any other two possible sugar ring conformations (Zhou et al., 1988). Elegant work with coupling constant data (Hosur et al., 1988) indicates little conformational averaging, and that a single pseudo-rotation angle intermediate between C2' and C3' *endo* is adequate to explain the experimental data for each residue.

The base-stacking arrangement is typical for B-type DNA (Figure IV-12). A greater overlap is found for purine-pyrimidine steps (steps 4-5, 7-8) than for pyrimidine-purine (steps 3-4, 6-7, 9-10), as is observed in the crystal structure of a DNA dodecamer (Dickerson & Drew, 1981), and as predicted by fibre diffraction studies (Arnott et al., 1969). The stacking arrangement appears symmetrical across the helix axis, as expected for B-DNA, and as seen in the dodecamer.

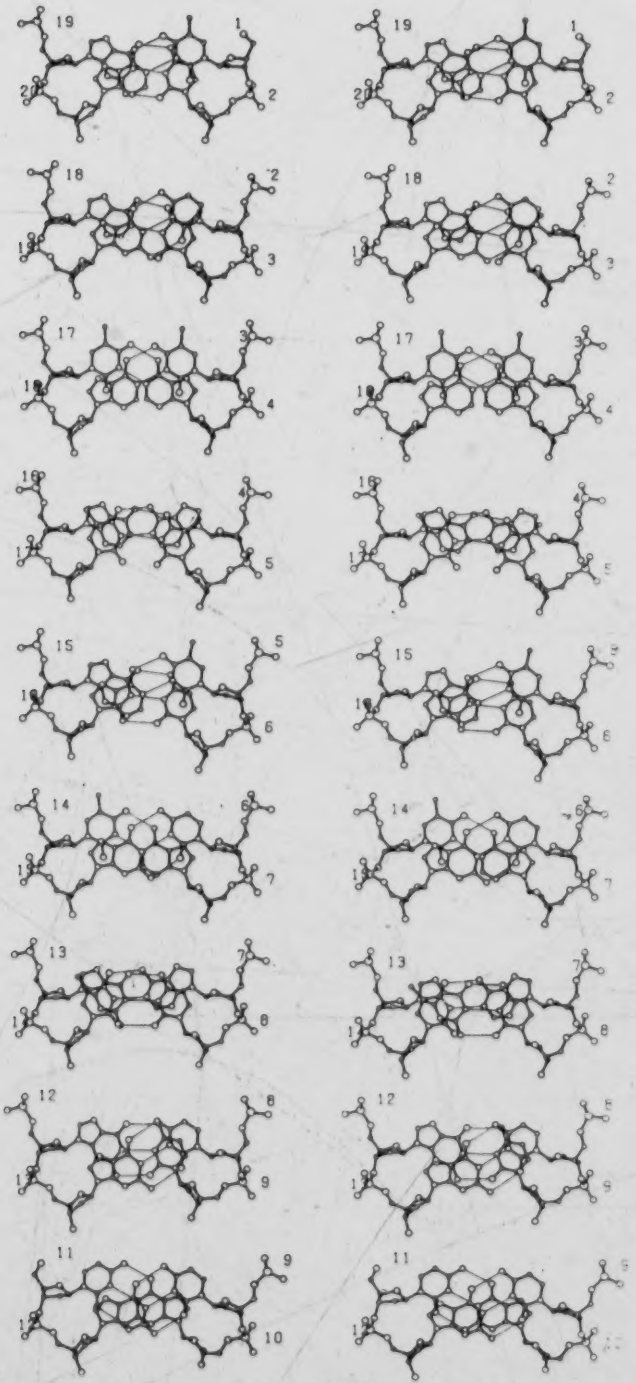


Figure IV-12 Base-stacking arrangements in the phage λ half-operator decamer
Views down the helix axis (perpendicular to that of Figure IV-8) are presented two base-pairs at a time.

D. Discussion

1. Accuracy of structure determination

There are a number of limitations on the accuracy of the structure determination presented here. Especially in the absence of a chosen molecular mechanics force field, it is still insufficiently understood to what extent the few hundred NOE intensities (or the derived short distances) and dihedral angles define, including a description of the dynamics, the three dimensional structure of oligonucleotides (van de Ven & Hilbers, 1988). For example, although the structures of the L10 decamer are experimentally determined to have phosphates in a B_I conformation, both starting models also have B_I phosphates, and there is no experimental data available that unambiguously distinguishes the two conformations, B_I and B_{II}, from one another.

Several lines of evidence however do support a B_I conformation. The positioning of the nucleotide units relative to each other, in the presence of even an approximate force field, is likely sufficient to localize the backbone torsion angles (Nilges et al., 1987). In crystal structures, the B_I conformation is more common than the B_{II} conformation, especially in regions of the DNA not affected by intermolecular hydrogen bonds or crystal contacts. This suggests that the B_I configuration is lower in energy and is more stable. Some B_{II} conformations are observed in B-DNA which have ε angles near -120°. This would correspond to a coupling constant between the 3' proton and the nearest phosphate greater than 10 Hz (Dickerson et al., 1985; Hosur et al., 1988). Figure IV-13 features an expansion of a typical non-terminal 2'<->3' cross-peak, with outer separations of the multiplet lines indicated in the 2' frequency dimension by $\Sigma J_{2'}$ and in the 3' dimension by $\Sigma J_{3'}$. These outer separations reflect the sum of coupling constants to the 2' proton ($\Sigma J_{2'} = J_{2'3'} + J_{2'2''} + J_{1'2'}$) and to the 3' proton ($\Sigma J_{3'} = J_{2'3'} + J_{2''3'} + J_{3'4'} + J_{3'P}$), with the active coupling, $J_{2'3'}$, giving the anti-phase characteristic to the cross-peak pattern. $\Sigma J_{3'}$ and $\Sigma J_{2'}$ are approximately 12 Hz and 30 Hz, respectively. For all sugar geometries, $J_{2'3'}$ is never less than 6 Hz (Hosur et al., 1988). From the absence of correlations in Figure IV-3, $J_{2''3'}$ is very near zero. However, all 3'-4' correlations are seen (Figure IV-6), implying a $J_{3'4'}$ of at least 2 Hz. Therefore, for all non-terminal residues, the 3'<->P coupling constant is small (< 5 Hz). From Figure 23 of the review by Hosur et al. (1988), the near 180° ε angles obtained here for a B_I conformation are consistent with this coupling constant being quite small.

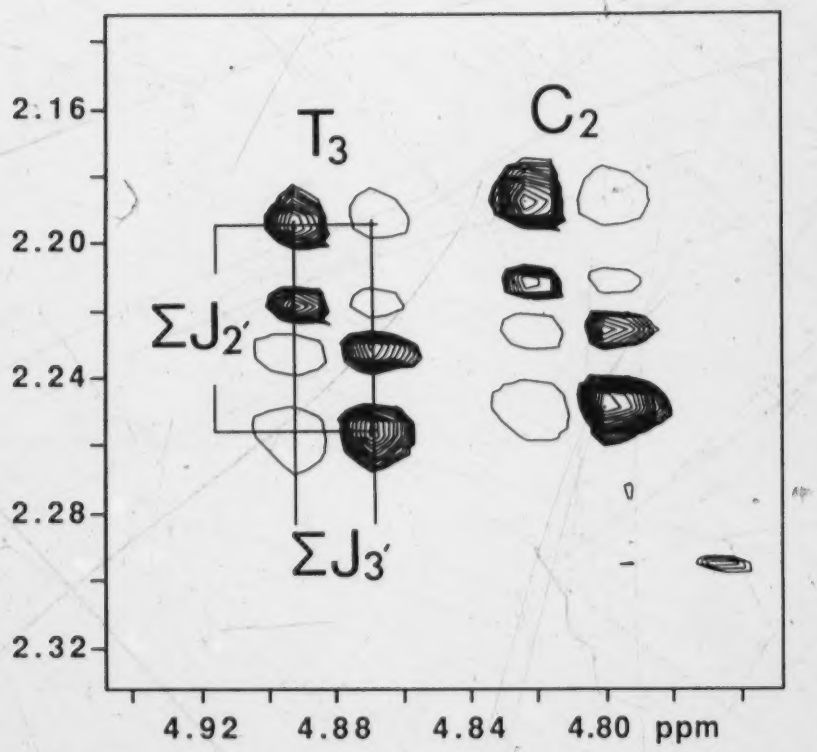


Figure IV-13 Expansion of 2' <-> 3' DQF-COSY crosspeaks for residues C_2 and T_3 $\Sigma J_{2'}$ and $\Sigma J_{3'}$ indicate the outer separations of the multiplet lines (approximately 30 and 12 Hz, respectively).

The MD calculation could start with a model including B_{II} phosphates, or more robust methods in conformational sampling could be employed, such as dynamic simulated annealing (e. g. Holak et al., 1988). The accuracy of the method described here appears to be between approximately 0.8 Å (from the rms deviation of structures R-A and R-B) and 1.8 Å (from the interconversion between the A and B-DNA fibre diffraction models). The 1.8 Å limit is potentially an over-estimate that would be improved by choosing better models which take sequence-dependent variations into account. The incorporation of NOE-based refinement circumvents inaccuracies in distances caused by neglecting indirect cross-relaxation pathways for the NOE between two protons and therefore one of the most serious problems in structure determination by NMR. Often at long mixing times, NOE cross-peaks can be observed between protons even 6 and 7 Å apart, but this information is lost since it cannot be directly related to a distance. The NOE refinement method indicates a way to obtain more parameters that cover a greater range of conformation space, and to better determine structures in solution by NMR.

As discussed in Chapter II, the initial motion model assumed that all interproton vectors tumble isotropically and are of invariant length. A more sophisticated model could be used, but the incorporation of model-free empirical parameters here is sufficient with respect to the precision of the NOE data experimentally collected.

2. Extension to the full operator structure

Figure IV-14 shows two structures for the full O_R3 operator DNA. On the left is the classical B-DNA conformation first used to model the interaction between Cro repressor and its operator DNA (Ohlendorf et al., 1982). Spheres in the major groove represent all atoms to which the protein is proposed to make sequence-specific contact (Takeda et al., 1989). Ethylation of the phosphates represented by cross-hatched spheres prevent Cro repressor binding (Ptashne et al., 1980). Other proposed non-specific contacts to phosphates are shown as small spheres. In Figure IV-14B, a consensus operator is built with two half-operators, using the conformation of the decamer determined above. The L10 sequence does not include the middle base-pair of the full operator. The central G•C base-pair is constructed here using average dihedral and twist angles and rise values between base-pairs. The similarity in the protein contact site between the models is striking, and suggests that the monomeric unit of Cro repressor recognizes DNA by the pattern of hydrogen-bonds and van der Waals contacts presented by the DNA, and not by any gross morphological feature.

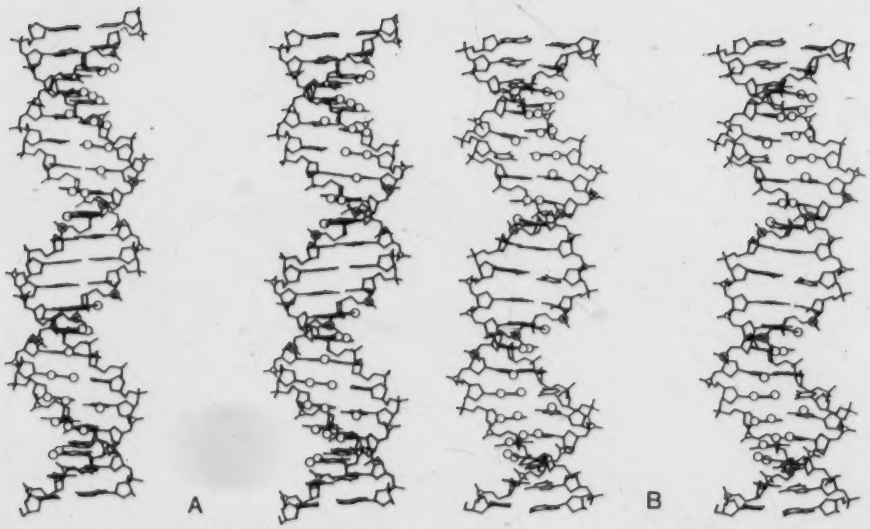


Figure IV-14 Models for the 21 base-pair O_R3 operator DNA

(A) B-type DNA. (B) The operator constructed using the solution structure of the half-operator (L10) determined using NMR, restrained molecular dynamics, and NOE-based refinement. Spheres in the major groove indicate the sequence specific contacts made by the Cro repressor protein. Ethylation of the phosphates denoted by cross-hatched spheres prevent Cro repressor binding. Other phosphate atoms proposed to be involved in non-specific protein contact are indicated by smaller spheres.

This study does not preclude any improvements in the model such as adjustments in the positioning of the two halves of the operator, for example, by bending at the centre. The significance of the difference between L10 and classical B-DNA awaits further improvements in the NMR methodology, and confirmation of the structure by other means, such as X-ray crystallography. A preliminary crystallographic determination of the Cro dimer-17 base-pair O_R3 complex (Brennan & Matthews, 1989b) indicates that, at the center of the operator, the DNA is overwound, and the minor groove is somewhat compressed[2]. A detailed comparison of the DNA structure in the Cro repressor sequence-specific site between free and protein-bound DNA is expected, and will be possible in the very near future.

3. Comparison between unbound DNA and in complex with the λ repressor

Recently, Jordan & Pabo (1988) have reported the crystal structure of the λ repressor DNA-binding N-terminal domain in complex with O_L1 operator DNA at a resolution of 2.5 Å (see Chapter I; Figure I-14). The consensus half of the O_L1 operator shares the same base sequence with base-pairs 3-10 of L10, enabling a comparison between unbound and bound conformation of the DNA. In Figure IV-15 the two structures are shown. Despite the use of different techniques for structure determination, the comparison is quite striking. The bound DNA has an overall atomic deviation of approximately 1.8 Å from classical B DNA and 1.3 Å from the corresponding eight base-pairs of L10. Widening of the major groove occurs at about base-pair 4 (second from the top of the figure) upon complexation with protein, probably for accomodation the recognition helix of the protein which allows the amino acid side-chains to make sequence-specific contacts to the edges of the base-pairs in this region. Widening of the minor groove and compression of the phosphates occurs at about base-pair 10 (bottom), in conjunction with changes near the center of the full operator. Phosphates across base-pair 5 also move closer togther. These changes reflect the alterations in DNA conformation upon binding λ repressor protein and similar alterations might be expected for the binding of Cro repressor protein.

---

[2]The Cro monomers retain nearly the same structure as observed in the crystal structure of the free protein (Anderson et al., 1981), although there are substantial changes in the relative positions of monomers in the Cro dimer.

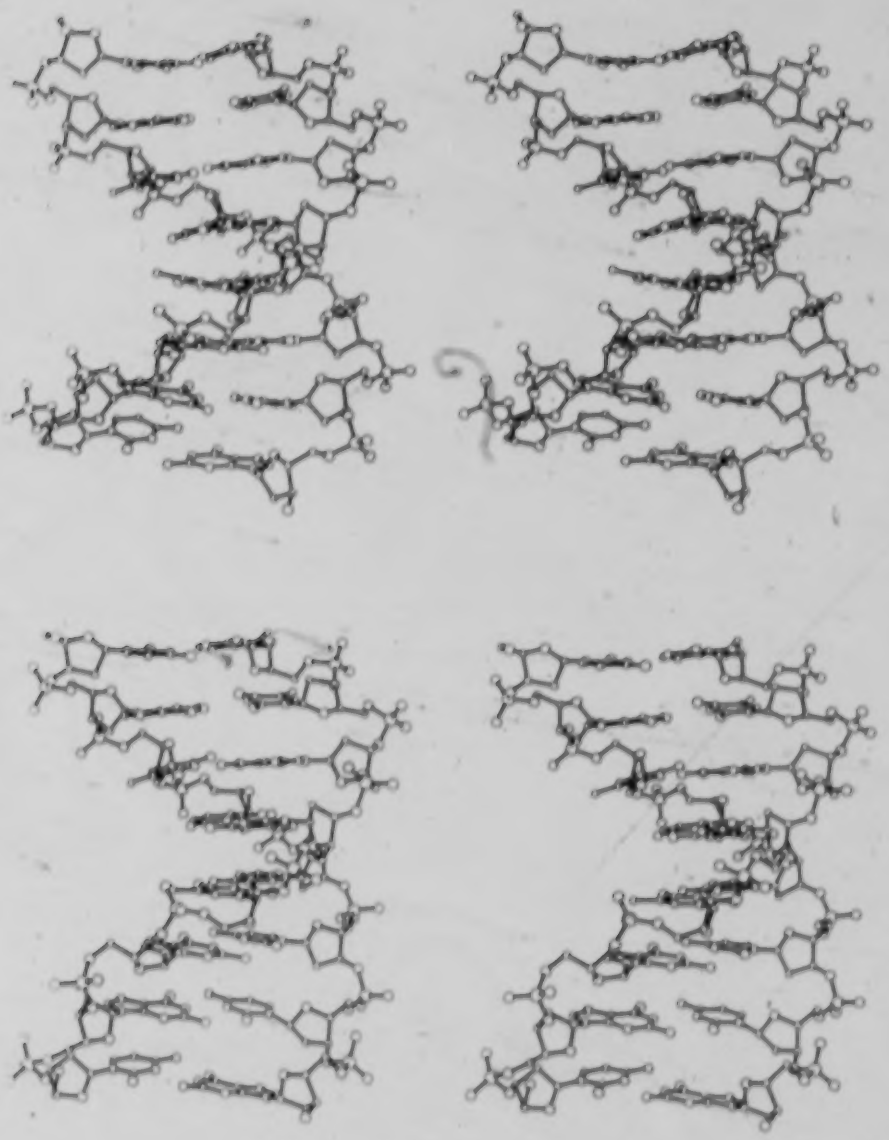


Figure IV-15 Structure of DNA free and in complex with λ repressor

Top: The consensus eight base-pairs derived from the L10 structure. Bottom: The corresponding base-pairs from the crystallographic determination of the λ repressor N-terminal domain-O_L1 DNA complex.

In Figure IV-16, an atom by atom comparison is shown. The two structures are superimposed and the difference in atom position is plotted against atom number, starting at the top left of the structures presented in Figure IV-15, continuing down the DNA strand, crossing to the other strand, and finishing at the top right. Especially in regions distant from protein contacts, the agreement is quite good. The largest deviations occur in what correspond to the center of the full operator, suggesting that significant alterations in the DNA backbone are concentrated here upon complexation with the protein, despite the absence of sequence-specific protein contacts at the center (as discussed in Chapter I).

The information content from NMR and crystallographic data is quite different and leads to different types of accuracy for the structures resulting from the application of the two techniques. Because there is a lack of NMR parameters directly relating to the DNA backbone configuration, the conformation about the phosphate is most poorly determined. In addition, while distances less than 5 Å are known to an accuracy of at least 0.5 Å, no longer range data is possible because the NOE is approximately proportional to the inverse sixth power of the distance between nuclei and therefore drops off rapidly for distances longer than 4 or 5 Å. X-ray data particularly is capable of placing the heavy atom phosphate atoms well, although the diffraction data contains information to a resolution of only 2.5 Å.

Figure IV-17 illustrates the some of the conformational parameters that are most dependent on base-sequence[3] (the roll angle and the helical twist angle) for the average refined structure of L10 and the eight corresponding base-pairs in the consensus half of the O_L1 operator bound to λ repressor protein. The roll angles do not agree well between bound and unbound DNA. The agreement for roll angle is good between the unbound DNA and predictions using the Calladine's (1982) rules (Figure IV-10). Since the Calladine prediction is accurate for the roll angle (see above, Chapter III; Dickerson et al., 1985), the differences for the bound DNA is either due to insufficient resolution or to induced changes by the protein. Better agreement exists for the helical twist angles between bound and free DNA, except near the center of the operator.

Part of these differences may arise from end effects, both for the NMR structure and the X-ray structure. The conformation parameters for the section of DNA that correspond to base-pairs 1 and 2 of L10 are not included because the base sequences are different,

---

[3] Propeller twist angles agree within 5°, except for the base-pair 9 (second from the bottom in Figure IV-15) where the unbound DNA has a propeller twist of 14° and the bound DNA has essentially zero propeller twist, probably due to induced conformational changes by the protein. Since the X-ray structure resolution was only 2.5 Å, no attempt was made to compare the difference in δ angles across the base-pair between bound and unbound DNA.

(possibly affecting the conformation of base-pair 3 as well), and crystal packing distortions were noted in X-ray structure (Jordan & Pabo, 1988). At base-pair 10 of L10, the lack of additional base-pairs, and additional mobility for this terminal base-pair may also result in discrepancies with the X-ray structure.

NMR has been used to determine the solution structure of a DNA sequence containing the half-operator consensus sequence from the regulatory operator of bacteriophage λ. This structure has been compared to the corresponding DNA in a protein-DNA complex obtained using crystallographic techniques. The agreement between the NMR and X-ray structures was quite good, especially in regions of the DNA not contacted by protein, showing that NMR can indeed be used to determine structures to high accuracy.

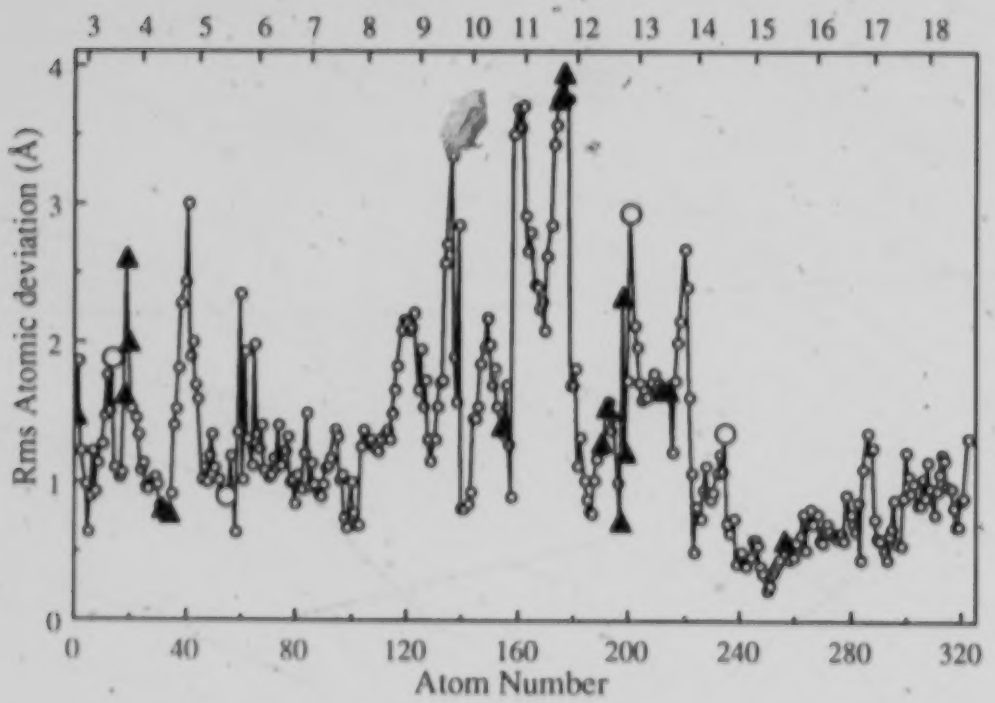


Figure IV-16 Atomic rms deviation between free and bound operator DNA
The structure of base-pairs 3-10 of L10 are compared with that of the corresponding base-pairs in the complex of O_L1 DNA with the λ repressor N-terminal domain (Jordan & Pabo, 1988). For reference, numbers indicate the C1' atoms of each residue. Hydrogen bonds made by protein are indicated by (▲), and van der Waals contacts by (o).

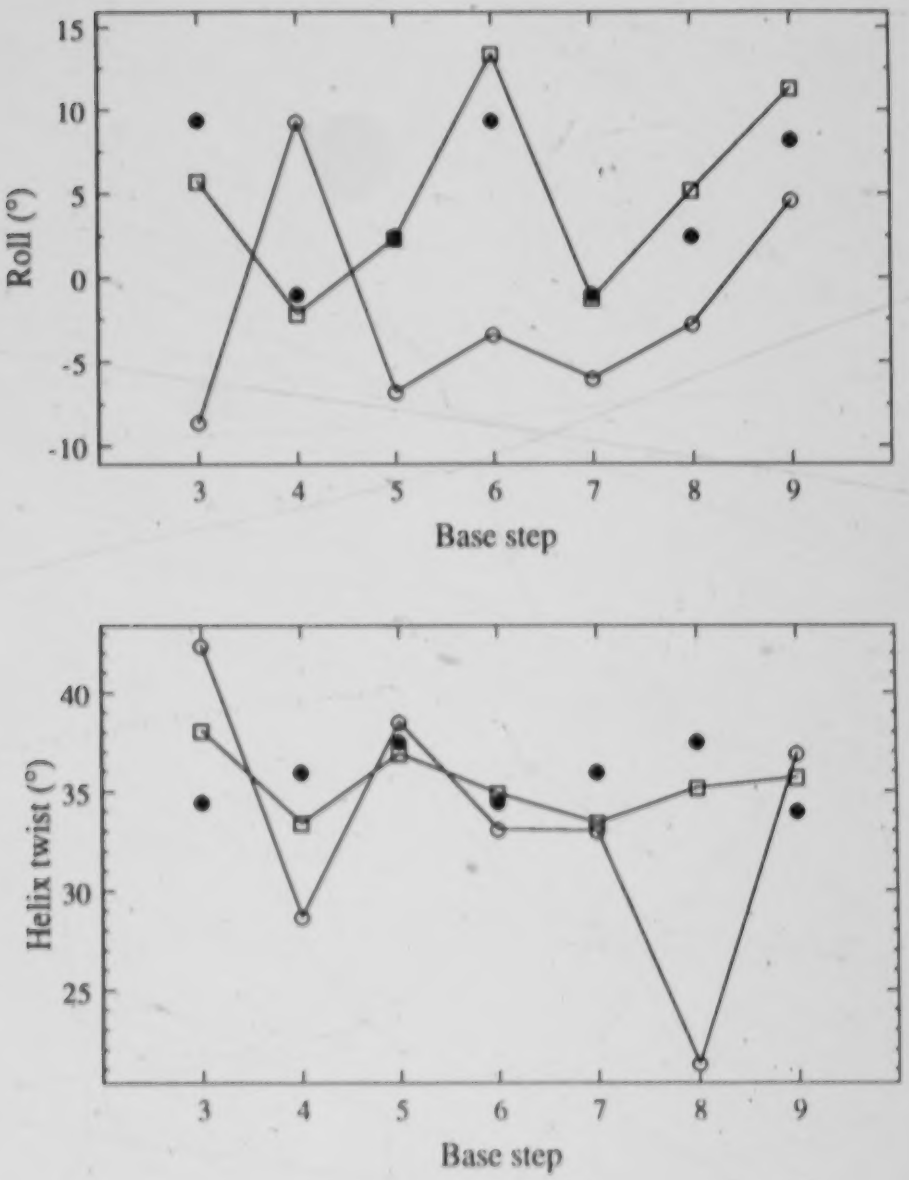


Figure IV-17 Comparison of conformation in free and complexed DNA

Variation in base roll and global helical twist for the L10 structure R_{ave} (-- □ --) and the corresponding base-pairs from the λ repressor N-terminal domain-operator DNA complex (—Δ—).

E. References

Aggarwal, A. K., Rodgers, D., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899-907.

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem Soc. 94*, 8205-8212.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754-758.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504-1509.

Arnott, S., Dover, S. D., Wonacott, A. J. (1969) *Acta Crystallogr. sect. B 25*, 2192-2206.

Baleja, J. D., Moult, J., & Sykes, B. D. (1990) *J. Magn. Reson.* (in press).

Brennan, R. G., & Matthews, B. W. (1989a) *J. Biol. Chem. 264*, 1903-1906.

Brennan, R. G., & Matthews, B. W. (1989b) *Trends Biochem. Sci. 14*, 286-290.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187-217.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343-352.

Chary, K. V. R., Hosur, R. V., Govil, G., Chen C., & Miles, H. T. (1988) *Biochemistry 27*, 3858-3867.

Cohen, J. S. (1987) *Trends Biochem. Sci. 12*, 133-135.

Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F., & Kennard, O. (1986) *J. Mol. Biol. 192*, 891-905.

Cuniasse, Ph., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., Guschlbauer, W., & Fazakerley, G. V. (1989) *Biochemistry 28*, 2018-2026.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Israel J. Chem. 27*, 181-188.

Dickerson R. E., Kopka, M. L., & Pjura, P. (1985) in *Biological Macromolecules and Assemblies, Volume 2, Nucleic Acids and Interactive Proteins*, (Jurnak, F. A. & McPherson, A., Eds.) pp 37-126, 490-493, Wiley-Interscience, New York.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419-441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761-786.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) *J. Biomol. Struct. Dynam. 5*, 557-579.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686-14707.

Gronenborn, A. M., & Clore, G. M. (1985) *Progr. Nucl. Magn. Reson. Spect. 17*, 1-33.

Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28*, 5978-5984.

Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Biophys. J. 16*, 279-286.

Hogan, M. E., & Jardetzky, O. (1980) *Biochemistry 19*, 3460-3468.

Holak, T. A., Engström, Å., Kraulis, P. J., Lindeberg, G., Bennich, H., Jones, T. A., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 7620-7629.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283-300.

Hosur, R. V., Ravikumar, M., Chary, K. V. R., Sheth, A., Govil, G., Zu-kun, T., & Miles, H. T. (1986) *FEBS Lett. 205*, 71-76.

Hosur, R. V., Govil, G., & Miles, H. T. (1988) *Mag. Reson. Chem. 26*, 927-944.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893-899.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179-182.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149-158.

Kirpichnikov, M. P., Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Kurochkin, A. V., Skryabin, K. G., & Bayev, A. A. (1984) *Nucl. Acids Res. 12*, 3551-3561.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1987) *Nature 326*, 886-888.

Koudelka, G. B., Harbury, P., Harrison, S. C., & Ptashne, M. (1988) *Proc. Natl. Acad. Sci. 85*, 4633-4637.

Lee, M., Chang, D.-K., Hartley, J. A., Pon, R. T., Krowicki, K., & Lown, J. W. (1988) *Biochemistry 27*, 445-455.

Lefèvre, J.-F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076-5090.

Lipari, G., & Szabo A. (1982) *J. Am. Chem. Soc. 104*, 4546-4559.

McKay, D. B., & Steitz, T. A. (1981) *Nature 290*, 744-749.

Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987) *Biochemistry 26*, 3734-3744.

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591-616.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455-475.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718-723.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187-193.

Otwinowski, Z., Schevitz, R. W., Zhang R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321-329.

Pabo, C. O., & Sauer, R. T. (1984) *Ann. Rev. Biochem. 53*, 293-321.

Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800-6801.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498-504.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M., & Sauer, R. T. (1980) *Cell 19*, 1-11.

Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dynam. 4*, 621-649.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327-337.

Scalfi Happ, C., Happ, E., Nilges, M., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 1735-1743.

Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom, J. H., Kaptein, R., Rüterjans, H., & Beyreuther, K. (1983) *Biochemistry 22*, 228-235.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.

Suzuki, E., Pattabiraman N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854-6865.

Takeda, Y., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1983) *Science 221*, 1020-1026.

Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. 86*, 439-443.

Tidor, B., Irikura, K., Brooks, B. R., & Karplus, M. (1983) *J. Biomol. Struct. Dynam. 1*, 231-252.

van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1-38.

van Gunsteren, W. F., Boelens, R., Kaptein, R., Scheek, R. M., & Zuiderweg, E. R. P. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., ed.) pp 92-99, Polycrystal Bookservice, Western Springs Illinois, U.S.A.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, Jr., S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765-784.

Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature 335*, 789-795.

Zhou, N., Manogaran, S., Zon, G., & James, T. L. (1988) *Biochemistry 27*, 6013-6020.

Chapter V

The Interaction of Phage λ O_R3 DNA with Wild type and V55C Cro Repressor Proteins

A. Introduction

In the preceding chapter, the structure of the left ten base-pairs of the O_R3 operator DNA sequence from phage λ, L10, was determined in solution using nuclear magnetic resonance techniques. In this chapter, the interaction of the half-operator with Cro repressor protein is studied in detail. The role of the Cro protein is to help regulate the life cycle of the temperate phage λ by binding strongly to a specific operator site, O_R3, on the viral DNA. In the absence of this protein, the virus can lie dormant within the bacterial host cell. Inducing events cause high Cro protein levels that result in activation of the lytic pathway where the virus is replicated and then lyses the host cell releasing phage progeny (reviewed in Chapter I; Ptashne et al., 1980). In order to function as a specific regulator of gene expression in transcription, Cro recognizes and binds preferentially to the O_R3 DNA sequence against the background of a large number of competing sites also present in the genome (Berg & von Hippel, 1988). In order to understand the details of the steric and physiochemical aspects of site-specific recognition process, nuclear magnetic resonance techniques are used to investigate the formation and structural features of Cro protein-DNA complexes in solution.

The left ten base-pairs of the 21 base-pair O_R3 operator sequence is chosen over the full operator because only the half-operator, L10, has sufficient resolution and narrow lines necessary for structure determination using NMR techniques (Chapter IV). The structure of the L10 half-operator free in solution and when bound by λ repressor protein is similar to classical B DNA (Arnott & Hukins, 1972), and suggests a direct read-out mechanism of DNA-protein recognition by the Cro repressor protein. The L10 sequence contains the consensus Cro repressor binding site, and should therefore have all the interactions observed between Cro and the full operator.

The L10 half-operator corresponds to the left ten base-pairs of the 21 base-pair O_R3 operator. Since the phage λ operator DNA sequences are nearly two-fold symmetric (pseudo-palindromic), two L10 molecules can simultaneously bind to Cro, with the second L10 taking the place of the right-half of the operator (Figure V-1).

This chapter gives experimental conclusions on: (1), the role of flexibility for Cro in forming the protein-DNA complex; (2), the binding strength and stoichiometry of the interaction between Cro repressor and L10 DNA; (3), the contacts made between protein and DNA; and (4), the inherent difficulties in using nuclear magnetic resonance techniques to study protein-DNA interaction.

```
O_R3    T C T A T C A C C G C A A G G G A T A A A
        A G A T A G T G G C G T T C C C T A T T T

L10     T C T A T C A C C G  C G G T G A T A G A   L10
        A G A T A G T G G C  G C C A C T A T C T
```

Figure V-1 O_R3 and L10 DNA sequences
Excepting the central base-pair, two L10 half-operators have nearly the same sequence as the O_R3 operator.

The life cycle of bacteriophage λ has long been a favorite for study in genetics and molecular biology. The λ Cro protein-operator DNA recognition system is no exception. The mechanism of action of the Cro protein began to be understood with its purification, sequence determination and assessment of nucleic acid binding properties (Folkmanis et al., 1976; Takeda et al., 1977; Roberts et al., 1977; Hsiang et al., 1977). The regulatory role of Cro protein in the life cycle of phage λ was seen to originate from its high affinity for the O_R3 DNA sites, as deduced by chemical and foot-printing experiments of the DNA (Johnson et al., 1978; reviewed by Ptashne et al., 1980; Shea & Ackers, 1985; Tullius & Dombroski, 1986; Hayes & Tullius, 1989) and by establishment of the levels and identity of the RNA transcripts regulated by Cro (Takeda, 1979). Crystallization of the protein (Anderson et al., 1979) and determination of its structure (Anderson et al., 1981) has since generated remarkable interest in the Cro repressor and its interaction with DNA.

Based on the crystal structure of the unbound protein, a model has been proposed for the sequence-specific interaction of Cro with DNA (discussed extensively in Chapter I; Ohlendorf et al., 1982; Anderson et al., 1983a; Ohlendorf & Matthews, 1983; Matthews et al., 1983). In the model, the amino acid side-chains of an α helix from the two monomeric units of Cro make contact to the edges of base-pairs exposed in two successive major grooves of B-type DNA. Sequence and structural homology in a helix-turn-helix motif with several phage DNA-binding proteins led to a search for similar substructures in other proteins (Anderson et al., 1982; Takeda et al., 1983; Ohlendorf et al., 1983a,b; Matthews et al., 1982; Pabo & Sauer, 1984; Cygler & Anderson, 1986; Dodd & Egan, 1987; Brennan & Matthews, 1989a). The Cro repressor protein has been characterized by florescence spectroscopy (Boschelli et al., 1982; Boschelli, 1982) and circular dichroism spectroscopy (Bolotina et al., 1983). One-dimensional ^{19}F and ^{1}H NMR studies have mainly investigated the aromatic residues of the protein (Arndt et al., 1983; Kirpichnikov et al., 1982a,b; Iwahashi et al., 1982; Kurochkin & Kirpichnikov, 1982; Kirpichnikov et al., 1984a,b; Kurochkin & Kirpichnikov, 1986). The ^{1}H NMR CIDNP effect in Cro repressor

has also been studied (Shirakawa et al., 1985a,b; Lee et al., 1987). ^{1}H NMR spectroscopy has been used to study a peptide fragment of Cro containing the recognition helix and its binding to DNA (Mayer et al., 1983). ^{1}H NMR assignments of the intact protein were essentially completed with two-dimensional techniques (Weber et al., 1985a,b) and an attempt[1] has been made to relate the structure of the Cro protein in solution to that of the crystal structure (Kurochkin et al., 1986). The utility of ^{15}N labelling in obtaining resonance assignments and understanding the dynamics of the Cro repressor was demonstrated by Leighton & Lu (1987).

Filter-binding assays have been extensively used to understand the energetics of binding in the Cro-DNA system (Kim et al., 1987; Sarai & Takeda, 1987; Takeda et al., 1989). The role of the lysine and tyrosine residues of Cro has been delineated by chemical modification studies (Takeda et al., 1986). A large number of site-specific mutants of Cro have been generated (Eisenbeis et al., 1985; Caruthers et al., 1986; Pakula et al., 1986; Hochschild et al., 1986; Hochschild & Ptashne, 1986; Caruthers et al., 1987; Pakula & Sauer, 1989; Takeda et al., 1989) to probe the details of the proposed model for the recognition of DNA by Cro repressor. Crystals of Cro repressor-DNA complexes have been obtained (Anderson et al., 1983b; Brennan et al., 1986) and a preliminary report of the structure of one of these complexes have been made (Brennan & Matthews, 1989b). As predicted by the Ohlendorf et al. (1982) model, the recognition α helices of Cro interact in two successive major grooves of a B-type DNA. The conformation of the Cro monomer is approximately the same in the complex as when it is unbound. However, the relative positions of the monomers change dramatically and the DNA bends, mainly at the center, to optimize the interaction between protein and DNA.

Studies centering on the DNA have not been ignored. CD spectroscopy has been used to show that the operator is in a B type DNA conformation, either free in solution, or bound to Cro protein (Kirpichnikov et al., 1985). The ^{1}H NMR resonances of unbound O_R3 DNA, and its analogues, have been assigned (Wemmer et al., 1985; Weiss et al., 1984; Hahn et al., 1985; Grütter et al., 1988). The imino protons (on N3 of thymine and N1 of guanine, Figure I-6) have been assigned for both the free DNA (Ulrich et al., 1983; Chou et al., 1983; Lee et al., 1983) and for the DNA-protein complex (Kirpichnikov et al.,

---

[1] In this paper a re-positioning of the α1 helix was postulated on the basis on a close approach of the α protons of Ile 5 and Phe 41. However, the α proton, and γ and δ methyl groups of Ile 5 were incorrectly assigned to the peaks at 4.9, -0.25, and .25 ppm, respectively (Figures V-5 and V-6). These peaks have been identified as belonging to the neighbouring residue Phe 41, Ile 40 (cf. Weber et al., 1985b; Lee et al., 1987).

1984a; Lee et al., 1987). Thymines of the O_R3 operator have been systematically replaced by fluorouracil, and ^{19}F NMR used to monitor the structure of DNA unbound, and in the presence of Cro repressor protein (Metzler et al., 1985; Metzler & Lu, 1989). In addition, systematic base-substitution experiments have been carried out on λ O_R1 DNA to assess the role of each base in the DNA for the affinity to Cro repressor (Takeda et al., 1989; Sarai & Takeda, 1987; Benson & Youderian, 1989).

The study of a protein-DNA complex using ^1H NMR spectroscopy is not an easy task. As will be shown below, two half-operators bind to the Cro dimer, creating a complex of approximately 28,000 in molecular weight. The proton resonances are broad and exhibit considerable spectral overlap. In addition, on forming the Cro-DNA complex, resonances often broaden beyond detection because of intermediate exchange behavior. The complexes are not often very soluble, and special experimental conditions must be employed to keep them in solution. Despite these limitations, data could be obtained from successful titrations of protein with DNA, and *vice versa* , and the results interpreted to give information on the role of flexibility of Cro and DNA on forming the protein-operator complex.

This chapter discusses two major experiments: (1), the comparison of the structure, dynamics, and DNA-binding characteristics of wild-type and a cross-linked Cro repressor, and (2), a more detailed study of DNA structure and DNA-protein interactions within the wild-type Cro DNA complex using ^1H NMR spectroscopy.

From the crystal structure of the Cro repressor (Anderson et al., 1981) valine 55 of one monomer is near to valine 55 of the other monomer (Figure V-2). In a mutant Cro protein, where valine 55 is replaced by cysteine (V55C), it will be shown by circular dichroism and ^1H NMR spectroscopies that the cross-link which forms between the subunits does not significantly perturb the structure from that of wild-type Cro in the absence of DNA at ambient temperatures[2] . However, the V55C Cro has lowered flexibility. The affinity of both proteins is examined for the L10 DNA half-operator sequence. The cross-linked protein has an eight-fold higher dissociation equilibrium constant than the native protein at 22°C, and suggests that adjustments in protein structure necessary to form a tight protein-

---

[2] The structures and dynamics of native and V55C Cro are only briefly compared. A more detailed comparison would require complete assignment and structure determination of the V55C repressor using NMR techniques. Possibly re-assignment and structure determination of the wild-type protein would be needed as well to confirm the assignments of Weber et al. (1985b), and for comparison to the X-ray structure (Anderson et al., 1981). To more fully characterize the stability of the proteins, a protein concentration study should be undertaken and the stability to other chemical denaturants could be measured. Consideration will be given for such experiments and for publication separate from this thesis.

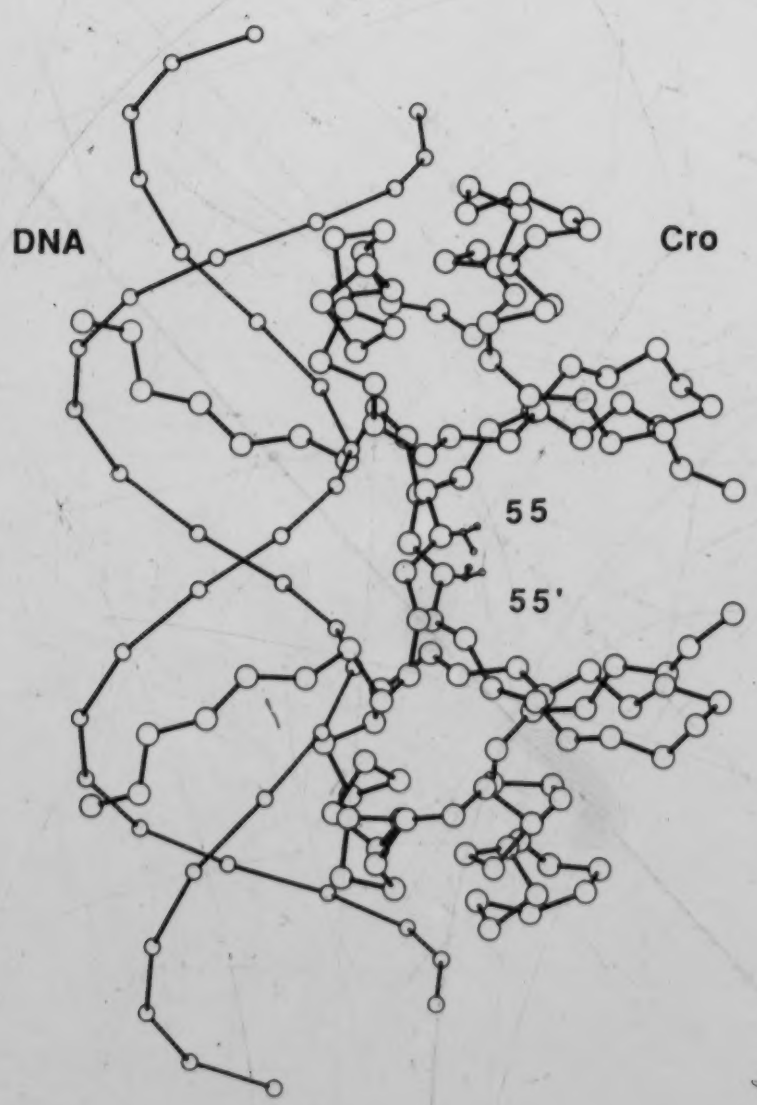


Figure V-2 The valine to cysteine substitution in Cro repressor
In the crystallographically determined structure of the wild-type Cro repressor, the side-chain of valine 55 of one subunit forms a close contact with valine 55' of the other subunit. This valine is replaced by cysteine in V55C Cro protein, resulting in a disulfide cross-link between the two monomers (Adapted from Figure I-8).

DNA complex are hindered by a loss in protein flexibility caused by the inter-subunit cross-link.

The second area is the study of the formation and DNA structure of the wild-type Cro-L10 DNA complex. Here, concentrations of protein and DNA are about 40 times higher. Non-exchangeable proton resonances of the DNA are monitored by ^{1}H NMR spectroscopy as wild-type Cro repressor was titrated in. Several protein resonances are assigned in the ^{1}H NMR spectrum of the protein-DNA complex and are used later (Chapter VI) to examine the conformation of the protein in Cro:DNA complexes.

B. Experimental procedures

1. Materials

The preparation of L10 DNA used in these experiments was as described in Chapter IV. Complex formation was also performed with the O_R3 17 base-pair operator (gift from Dr. W. Anderson). Wild-type and V55C Cro protein sample purification began by obtaining plasmids encoding the proteins under the control of isopropylthiogalactoside (IPTG) inducible transcription (from M. H. Caruthers and W. Anderson). Plasmids were incorporated into host *E. coli* by the $CaCl_2$ transformation procedure and stored in 20% glycerol until needed (Maniatis et al., 1982). Plasmids pTR214 and pJS303, which carry the gene for wild-type Cro, were transformed into *E. coli* host strains K802 and MV1190, respectively (Roberts et al., 1979). The valine 55 to cysteine substituted Cro was on a plasmid termed pV55C, and was transformed into host strain RB791. All plasmids carried ampicillin resistance. Plasmid/*E. coli* systems produce Lac repressor which binds to a site engineered upstream of *cro*, and prevent *cro* transcription in the absence of the inducer IPTG or lactose. Adding an inducing agent to the medium reduces the affinity of Lac repressor for DNA and permits *cro* transcription to proceed (Eisenbeis et al., 1985; Caruthers et al., 1986).

Prior to growing large amounts of transformed cells, it was demonstrated on smaller culture that the cells could be induced. Transformed *E. coli* cells occasionally did not produce Cro repressor, but still maintained ampicillin resistance. Individual colonies of transformed cells were replicate-plated onto LB agar plates (Maniatis et al., 1982) with and without 4 mM IPTG. Colonies of cells producing large amounts of Cro failed to thrive on IPTG containing plates, and the corresponding, or replicate colonies on the other plate could be used as a stock culture for a large scale preparation.

Alternatively, 10 mL liquid cultures (containing 10 g/L Bacto-tryptone, 5 g/L yeast extract, 8 g/L NaCl, pH adjusted to 7.2 with NaOH, and 50 mg/L ampicillin) were inoculated with a colony of the transformed cells, grown to an optical density at 600 nm. (OD_{600}) of 0.7 at 37°C, and induced with 0.5 mM IPTG for 3 hours. The optical density of cultures producing Cro was approximately 2.0, and cultures not producing large amounts of Cro had OD_{600} values near 3.0. Cro production was also checked by denaturing polyacrylamide gel electrophoresis using a Laemmli buffer system (Laemmli, 1970), for which Cro shows as the smallest protein when stained with Coomassie Blue. The cells from 0.5 mL OD_{600} of culture (0.5 mL of a culture of OD_{600} = 1.0) were centrifuged and taken up in 100 µL of 2 M urea, 2% SDS, 15 mg TRIS, pH 6.8, 12% glycerol, 0.01% bromphenol blue, and boiled for two to three minutes. Typically 15 µL were applied to a gel slot. After electrophoresis and Coomassie blue staining, gels were scanned with a densitometer. Of the total stained cell protein, typically <0.1 to 1.0% was measured to be wild-type Cro in the induced pTR214/K802 system, 1 to 10% of the total was wild-type Cro in the pJS303/MV1190 system, and 5-25% of the total cellular protein was V55C Cro protein in the induced pV55C/RB791 system.

A typical preparation of Cro protein used 10 L of culture (evenly distributed between ten 4 L flasks). After 3 hours of induction with 0.5 mM IPTG (as above) cells were harvested by cooling on ice and centrifugation (4500 rpm for 15 minutes). Cells were frozen (approximately 25 g) and stored for a few days at -20°C. Subsequent steps were done at 4°C. Clumps of frozen cells were homogenized and 150 mL of 10% sucrose, 50 mM TRIS, pH 7.4, 0.1 mM EDTA and 50 mM KCl were added. 20 mg of lysozyme (hen egg white) was added and the solution incubated for 30 minutes on ice. The temperature was then raised to 30°C briefly, and then reduced again on ice. 0.15 mL of 1 M $MgCl_2$ was added, and approximately 0.25 mg of pancreatic DNaseI. To ensure complete lysis, the cell purée was sometimes sonicated for several pulses of about 30 seconds. 3.25 g of KCl were dissolved and the suspension was centrifuged (40,000 rpm for 1 to 2 hours) to remove cellular debris[3] . The supernatant was diluted with 10 mM KH_2PO_4, pH 6.4, 0.3 mM EDTA, 5% glycerol buffer by approximately 3.5 times and applied to a 1.5 by 60 cm. phosphocellulose column (Whatman P-11). Cro was then purified by phosphocellulose and Sephadex G75 chromatography (Folkmanis et al., 1976). Further purification of Cro (from that produced using the pTR214 plasmid) was carried out by adding $(NH_4)_2SO_4$ to the pooled and concentrated peak from gel exclusion chromatography (approximately 3 mg/mL

[3] From analysis on denaturing gels, some Cro was present in the pellet. Raising the pH to about 8.0 prior to centrifugation minimizes the losses at this step (G. Ozimek, personal communication).

protein) to a final concentration of 40% (saturation), and adding saturated $(NH_4)_2SO_4$ dropwise over several says and growing microcrystals (Anderson et al., 1979). Gel electrophoresis, UV spectroscopy, and ^1H NMR spectroscopy confirmed product homogeneity to be greater than 95%. V55C Cro protein behaved the same as wild-type protein during purification, except that it ran on a denaturing gel at twice the molecular weight of the wild-type protein in the absence of β-mercaptoethanol. Protein concentrations were determined by amino acid analysis and by UV absorption spectroscopy using a molar extinction coefficient of 3970 (Boschelli, 1982).

Unless otherwise stated, all protein and DNA samples were made up in a standard buffer of 0.2 M KCl, 10 mM K_2HPO4, 10 mM KH_2PO4, 50 μM EDTA, pH 7.1±0.1. Solvent exchange was carried out by lyophilization procedures. All protein concentrations refer to the dimer form of the repressor. All circular dichroism studies were carried out in H_2O; NMR studies were carried out in D_2O, unless noted otherwise.

2. CD Spectroscopy

All protein and DNA samples were dialyzed (separately) against 0.2 M KCl, 10 mM K_2HPO4, 10 mM KH_2PO4, 50μM EDTA, at pH 7.3 and filtered before use. All subsequent sample manipulations and recording of CD spectra were performed by K. Oikawa. The CD measurements were made on a Jasco J-500C spectropolarimeter with a DP500N data processor, equipped with a thermostated cell holder (as described in Williams et al., 1986). Far-UV studies were made using a 1 cm cell (600 μL volume) with the temperature inside the cell calibrated with a Fluke multimeter with an 80 T150 temperature accessory. The mean residue molar (or molecular) ellipticities were calculated (Oikawa et al., 1968) with a mean residue weight for λ Cro of 111.6. The band intensity at 222 nm. was monitored for thermal denaturation studies of the Cro proteins.

3. NMR Spectroscopy

All NMR spectra were obtained on a VXR-500 NMR spectrometer with an operating frequency of 500 MHz for protons. One-dimensional spectra in 85% H_2O/ 15% D_2O were taken at 22 and 35°, with a 1-1 binomial suppression of the H_2O peak (Chapter IV). Typically 1024 transients with an acquisition time of 0.5 seconds and a relaxation delay time of 2.0 seconds were summed into 13K data points. A line-broadening of 5 Hz and apodization function time constant, AF, (Chapter III) of 0.2 seconds were applied for processing of the spectra. One-dimensional spectra at millimolar concentrations of protein

and DNA in D_2O were typically obtained by the sum 16 to 64 transients, which were each acquired in 0.5 seconds with a relaxation delay between scans of 1.8 seconds, and were processed with a line-broadening of 0.5 Hz. Spectra of proteins at lower concentrations (near 50 μM) were obtained with 1024 transients and with an applied line-broadening of 2 Hz. Two-dimensional spectra in D_2O were taken as described in Chapter IV. The mixing time for the proteins in the absence of DNA was 0.150 seconds. NOESY spectra of the wild type Cro:L10 DNA complex were taken with a mixing time of 0.250 seconds.

4. Binding assays

The binding of two L10 DNA molecules, L, to the Cro protein macromolecule, P, can be described by the molecular step-wise equilibrium:

$$P + L \underset{}{\overset{K_1}{\rightleftharpoons}} LP \overset{K_2}{\rightleftharpoons} PL_2 \qquad (1)$$

with $$K_1 = \frac{[P][L]}{[LP]} \qquad (2)$$

and $$K_2 = \frac{[LP][L]}{[PL_2]} \qquad (3)$$

The titrations presented in this chapter monitor a change in a signal of the DNA (chemical shift or line width) as Cro repressor protein is added. Let the signal of DNA in the absence of protein be δ_f. The signal of the DNA when bound in a LP complex is δ_{b1} and when bound in a PL_2 complex, δ_{b2}. The total concentration of protein is represented by:

$$P_o = [P] + [LP] + [PL_2] \qquad (4)$$

and that of DNA by:

$$L_o = [L] + [LP] + 2[PL_2] \qquad (5)$$

The signal is in fast exchange, meaning the average of free and bound signals are observed, so that δ_{obs} is $f_f \delta_f + f_{b1} \delta_{b1} + f_{b1} \delta_{b1}$ where f_f, f_{b1}, f_{b2} are the fractions of DNA free and bound, respectively ($f_f + f_{b1} + f_{b2} = 1$). The observed change in the signal from δ_f, $\Delta\delta$, is:

$$\Delta\delta = \delta_{obs} - \delta_f = \frac{[LP]}{L_o} \delta_{max1} + \frac{2[PL_2]}{L_o} \delta_{max2} \qquad (6)$$

where δ_{max1} is δ_{b1}-δ_f and δ_{max2} is δ_{b2}-δ_f. Known values are $\Delta\delta$, P_o, and L_o. Equations 4 to 8 are solved with a cubic function[4] and the interrelated K_1, K_2, δ_{max1} and δ_{max2} values are fit with a non-linear least squares method (Hull, 1975). Possible subsequent manipulations of the data have been discussed by Ferguson-Miller & Koppenol (1981). The binding of one DNA molecule to the Cro dimer is discussed in Chapter VI.

On the binding-site level, the interaction of Cro protein and L10 DNA is described by microscopic binding equilibria:

$$\begin{array}{ccccc} & \beta_1 & \text{PL} & c\beta_2 & \\ & \rightleftharpoons & & \rightleftharpoons & \\ \text{P} & & & & \text{LPL} \\ & \rightleftharpoons & & \rightleftharpoons & \\ & \beta_2 & \text{LP} & c\beta_1 & \end{array} \qquad (7)$$

where β_1 and β_2 are the microscopic site dissociation binding constants, and c is a cooperativity factor that indicates how the presence of one DNA already on the Cro dimer affects the binding of a second L10 DNA molecule. If c is equal to 1, the two sites are non-interacting. If c is less than 1, the dissociation binding constant for the addition of a second L10 is smaller, meaning enhanced binding (and *vice versa*). Since all the species in equation 7 can be considered to have full two-fold symmetry, LP=PL, and $\beta_1 = \beta_2 = \beta$ with:

$$\beta = 2\frac{[P][L]}{[PL]} \qquad \text{and} \qquad c\beta = \frac{1}{2}\frac{[PL][L]}{[LPL]} \qquad (8)$$

where the factors of 2 indicate that there are two sites on the protein for association of the LP complex and that the LPL complex can be formed from either LP or PL complexes (Ferguson-Miller & Koppenol, 1981). Therefore,

$$K_1 = \beta/2 \qquad \text{and} \qquad K_2 = 2c\beta \qquad (9)$$

The binding strengths for the formation of Cro-DNA complexes are subsequently referred to in terms of the cooperativity factor, c, and a microscopic equilibrium dissociation binding constant, K_d, defined to be equal to β.

---

[4] Handbook of Chemistry and Physics (Hodgman, C. D., ed.), 44th edition, p. 320, Chem. Rubber Pub. Co., Cleveland, Ohio.

C. Comparison of wild-type and V55C Cro proteins

1. Similarities in conformation and dynamics

The one-dimensional ^1H NMR spectra of the native and cross-linked Cro proteins are shown in Figure V-3, with selected regions expanded in Figures V-4 to V-6. There is generally a close correspondence in chemical shifts between the proteins. Partial assignments to specific residues were made for V55C protein by following the assignment procedures outlined by Wüthrich (1986) with some guidance from the published assignments of the wild-type protein (Weber et al., 1985b; see also Footnote 2 of this chapter). In Figures V-3 to V-6 several well-resolved lines are numbered with the assignment to the amino acid residue. Some of these resonance lines are used later to follow the alteration in the conformation of the protein with the addition of DNA.

The one-dimensional spectra highlight the overall similarity between the two proteins. Resonances of the aromatic side-chains (Figure V-4), have the same chemical shift positions for residues that are in α-helices (such as Tyr 10, Phe 14, Tyr 26, and His 35). In particular, the amino acid side-chains on the recognition helix (Tyr 26 and His 35) have unaltered chemical shifts, indicating that DNA-binding should still be possible. Amino acid side-chains belonging to residues near the cross-link site, such as Tyr 51 and Phe 58, have changed slightly in resonance positions, indicating small adjustments in this region of the protein in order to accommodate the disulfide bond.

The region of the ^1H NMR spectrum (Figure V-5) immediately downfield of the residual HOD shows resonances of α protons of residues in a β-sheet conformation (Wüthrich, 1986). Some resonance positions do not change (Ala 52, Glu 54, Leu 42, Ile 40) while others change by small amounts (about 0.1 ppm). Much of the shift observed for the α proton for residue 55 can be attributed to the replacement of the valine methyl groups by sulfur (Bundi & Wüthrich, 1979).

The methyl group resonances of the two proteins are compared in Figure V-6. The upfield shifted resonances of Ile 40 and Leu 42, in the hydrophobic core of the protein, and the methyl groups of Thr 64, Thr 65 and Ala 66 belonging to the flexible C-terminal arm of the protein have not appreciably changed resonance positions. The thiomethyl group resonance of Met 12 has not changed chemical shift, although the SCH_3 group of Met 1 has. The thiomethyl group of the wild-type protein was assigned by pH titration where the ionization of the N-terminal amino group (pKa = 7.7) exerts a weak but fully identifiable influence on the chemical shift of the SCH_3 group (Kurochkin & Kirpichnikov, 1986; unpublished experiments). Some of the largest chemical shift changes occur for

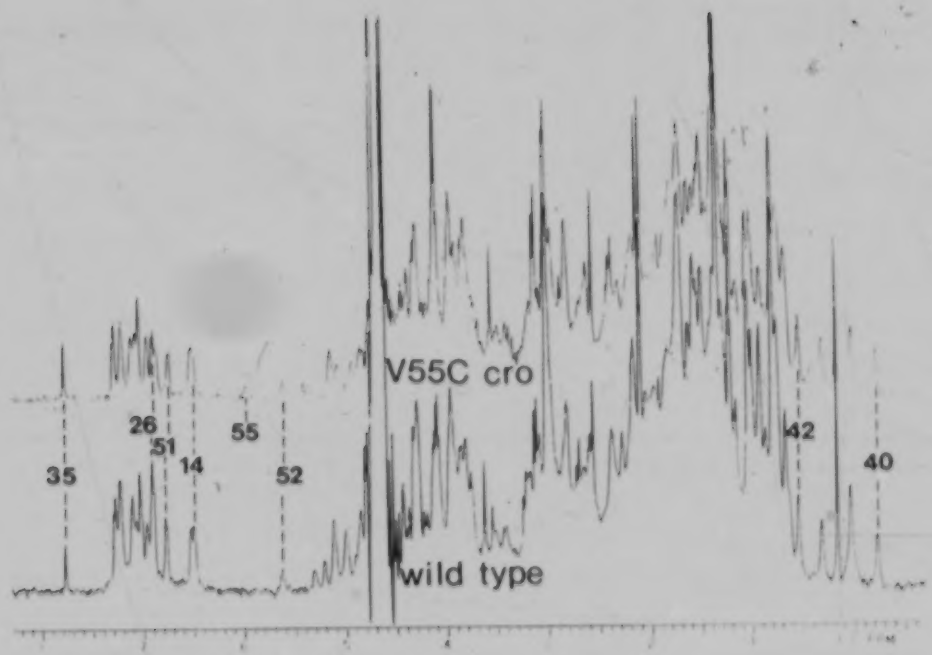


Figure V-3 Comparison of ^{1}H NMR spectra for native and V55C Cro protein
^{1}H NMR spectra of 40 μM protein were taken in standard buffer at 35°C. The numbers indicate assignment to particular residues of the proteins.

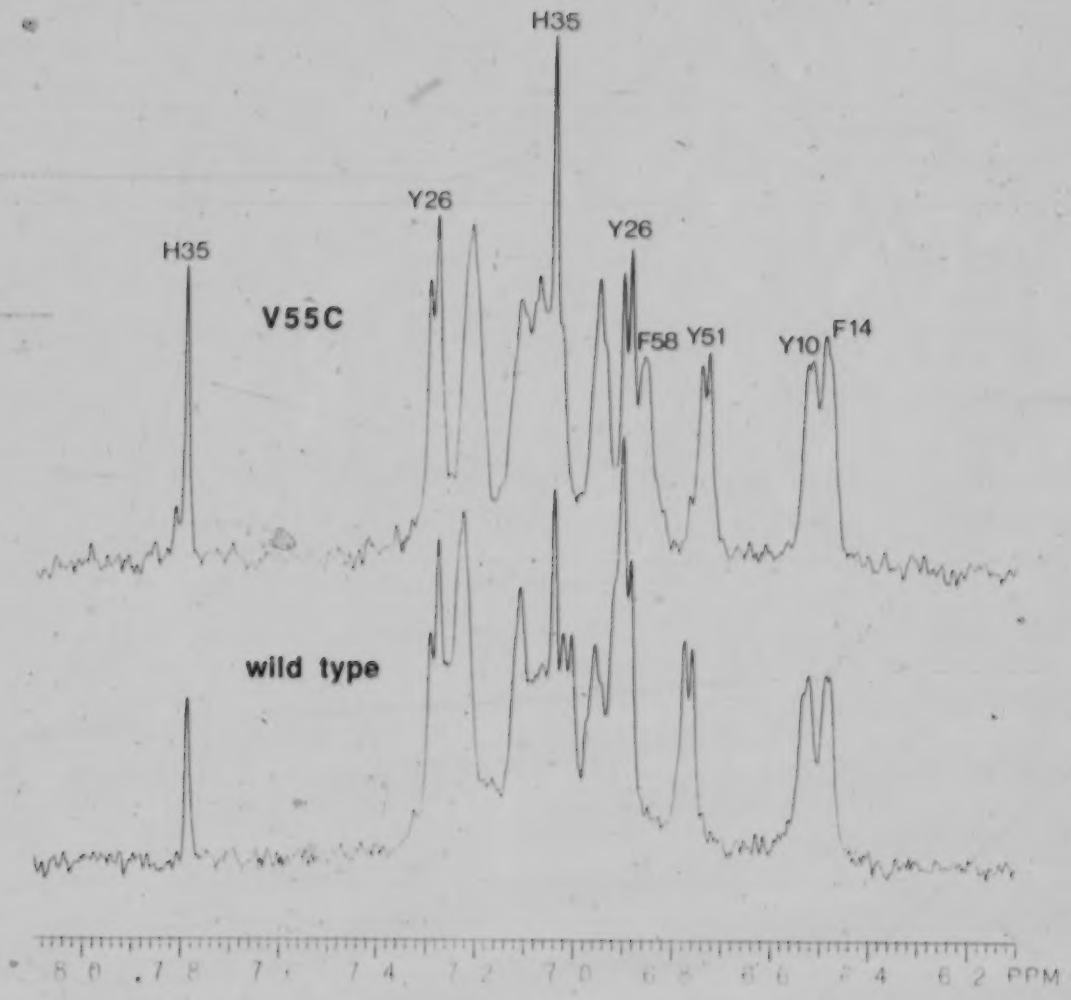


Figure V-4 ^{1}H NMR spectra of native and V55C Cro aromatic side-chains
^{1}H NMR spectra of 40 μM protein were taken in standard buffer at 30°C. The numbers indicate assignment to particular residues of the proteins.

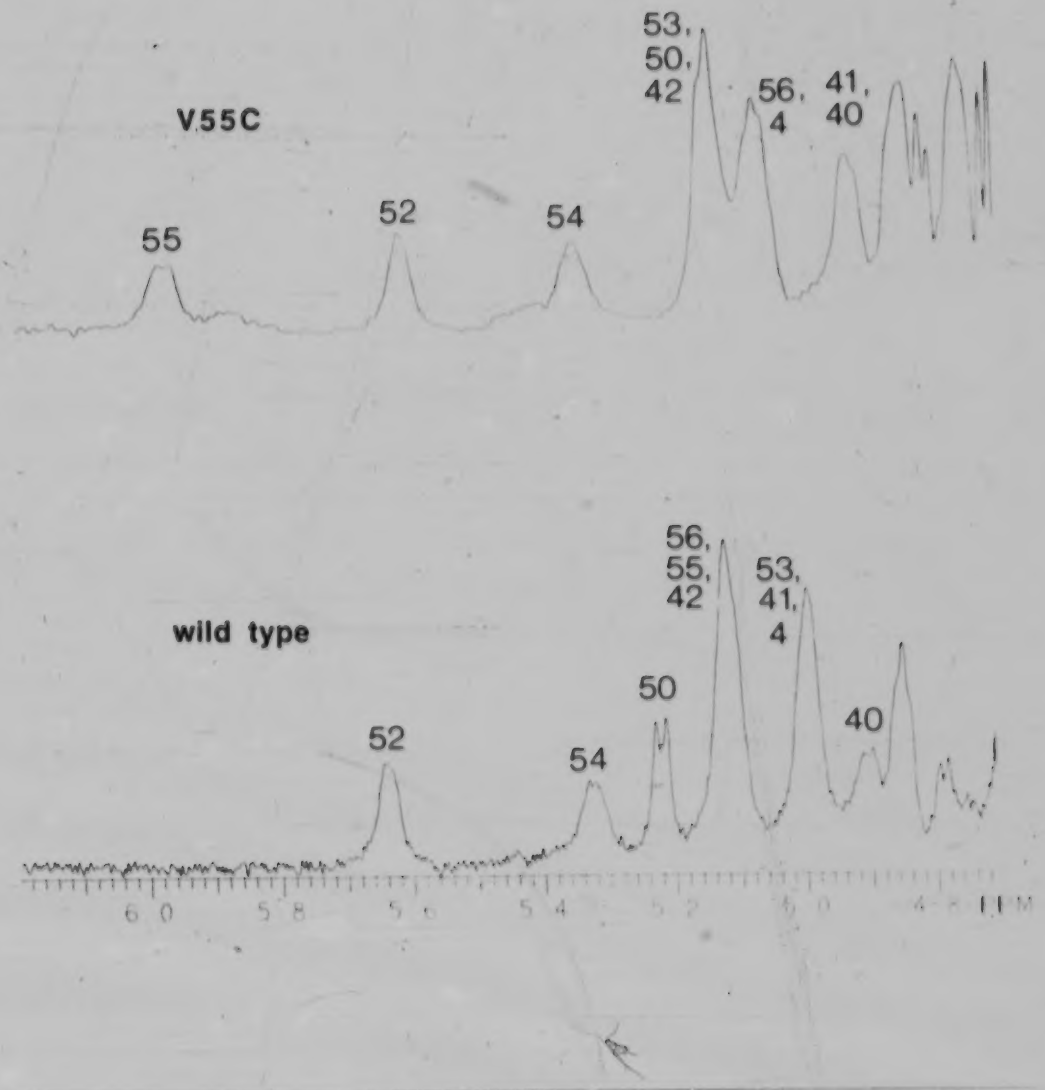


Figure V-5 ^{1}H NMR spectra of native and V55C Cro α protons

^{1}H NMR spectra of 0.75 mM wild type Cro protein were taken in 50 mM KCl, 5 mM K_2HPO_4, 5 mM KH_2PO_4, pH 6.9, 75 μM EDTA, at 30°C. V55C Cro was 1.5 mM in concentration is 100 mM KCl, 5 mM K_2HPO_4, 5 mM KH_2PO_4, pH 6.0, 75 μM EDTA, at 30°C. The spectra were processed with 1 Hz line broadening. The numbers indicate assignment to particular residues of the proteins.

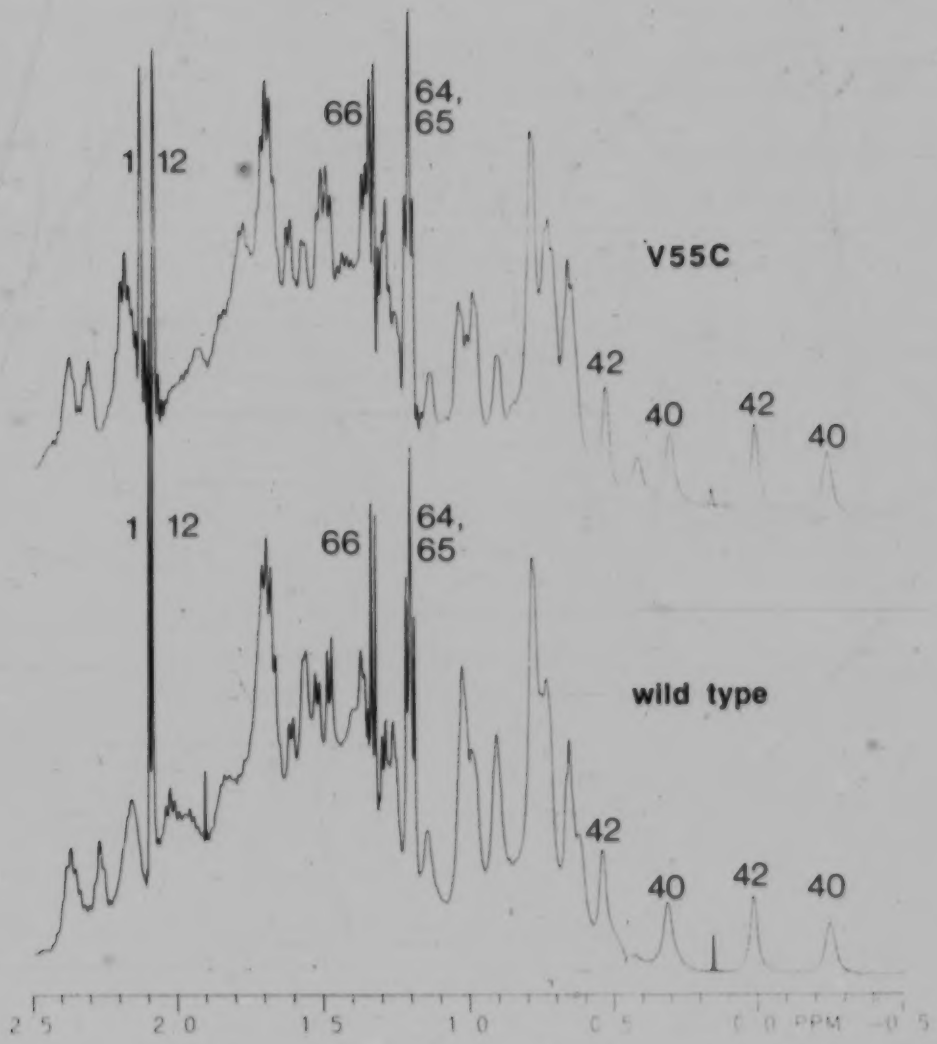


Figure V-6 ^{1}H NMR spectra of native and V55C Cro methyl protons

^{1}H NMR spectra of the wild type and V55C Cro proteins were taken at 30°C. Sample conditions are given in Figure V-5. The numbers indicate assignment to particular residues of the proteins.

residues, such as Met 1 and Phe 41, at the portion of the β sheet distal to the site of replacement. An alteration is possible between the monomer domains in the dimer interface or hydrophobic packing region at this end of the protein.

The NOESY experiment has been used to determine and compare spatial relationships of the protons in the native and V55C mutant Cro repressors. To examine in more detail the adjustments in structure near the cross-link site, the NOEs between the β-strands of the proteins are compared in particular. In an anti-parallel β sheet, the α proton of one strand is close (approximately 2.3 Å) to the α proton of the other hydrogen-bonded strand (Wüthrich, 1986; Figure V-7). In the corresponding NOESY spectra (Figure V-8) cross-peaks are observed between the resonance frequencies of α-protons where these protons are close in space. The same α proton to α proton connectivities are observed for the two proteins with each cross-peak being of approximately equal intensity in both spectra. Also observed in this spectral region are cross-peaks indicating a Ser 6 β to Phe 41 α contact, and the close proximity of the α proton for Lys 56 to the δ protons for Pro 57. The similarity in cross-peaks indicate the same interproton distances and therefore a close correspondence of the core β sheet structure of the two proteins.

The normally protonated V55C Cro protein was re-dissolved in deuterium oxide (D_2O). After approximately one day at 35°C, the amide protons for six residues are not exchanged with solvent deuterons. In a COSY type experiment, these remaining amide proton resonances give rise to cross-peaks with α protons (Figure V-9). Protons resistant to exchange result from shielding from solvent, usually by being involved in extensive hydrogen-bonding (Englander & Kallenbach, 1979). All six protons map to residues that correspond to the central β sheet core of wild-type Cro. In Figure V-10, a qualitative structure for the V55C Cro protein is constructed using the known amino acid sequence, the observed NOE connectivities between α protons, and incorporation of hydrogen-bonds to the six slowest exchanging amide protons. This structure is consistent with that observed in the crystal structure of the wild-type Cro repressor (Anderson et al., 1981).

Circular dichroism spectra (Figure V-11) also highlight the similarities between the two proteins. CD spectra represent the differential absorption of left and right circularly polarized light, and can be interpreted in terms of contributions from protein secondary structural elements, such as α helix, β-sheet and random coil (Cantor & Schimmel, 1980). There are some differences around 200 nm., where the magnitude of the molar ellipticity is greater for the cross-linked protein, and around 215 nm., where it is slightly less. The magnitudes of these differences are small, suggesting a change in conformation for not more than one or two amino acid residues.

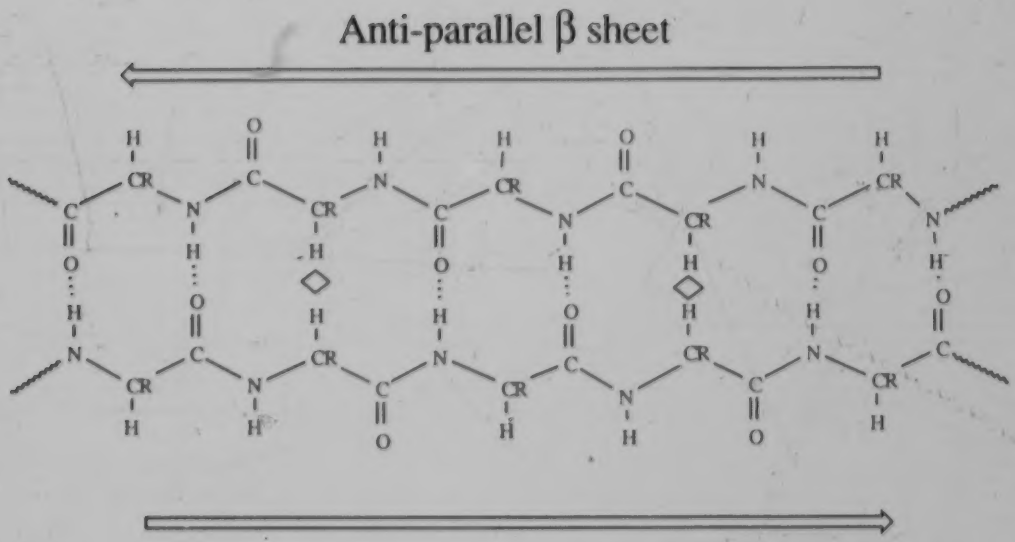


Figure V-7 Antiparallel β sheet structure

In an antiparallel β sheet, the α proton of one strand is close to the α proton of the other hydrogen-bonded strand (indicated by ◇). If a contact exists between the α proton of residue i for one strand and residue j of the neighboring strand, the next short contact occurs two residues further along the chain in an antiparallel fashion, e. g. at i+2 and j-2.

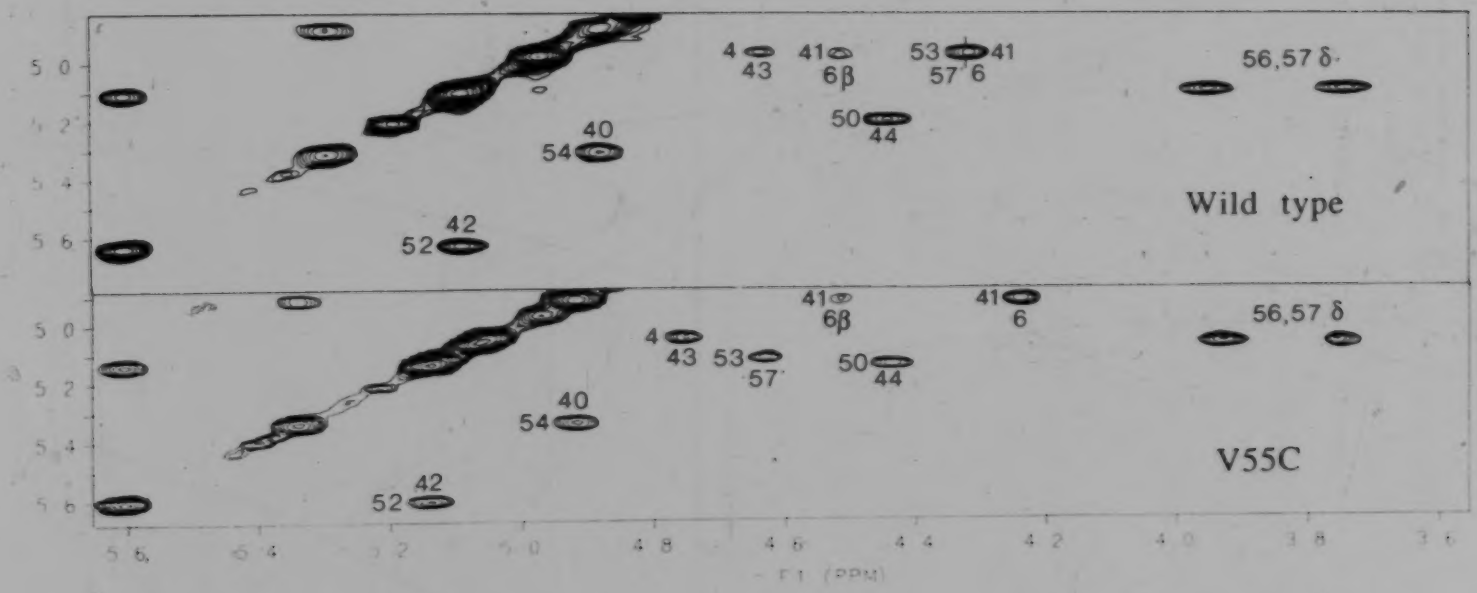


Figure V-8 NOESY spectra of wild type and V55C Cro proteins

This region of the two-dimensional NOE spectrum shows NOE cross-peaks between α protons upfield of the residual HOD. These indicate short contacts between β strands in the antiparallel β sheet of each protein. Numbers indicate the assignment to specific residues.

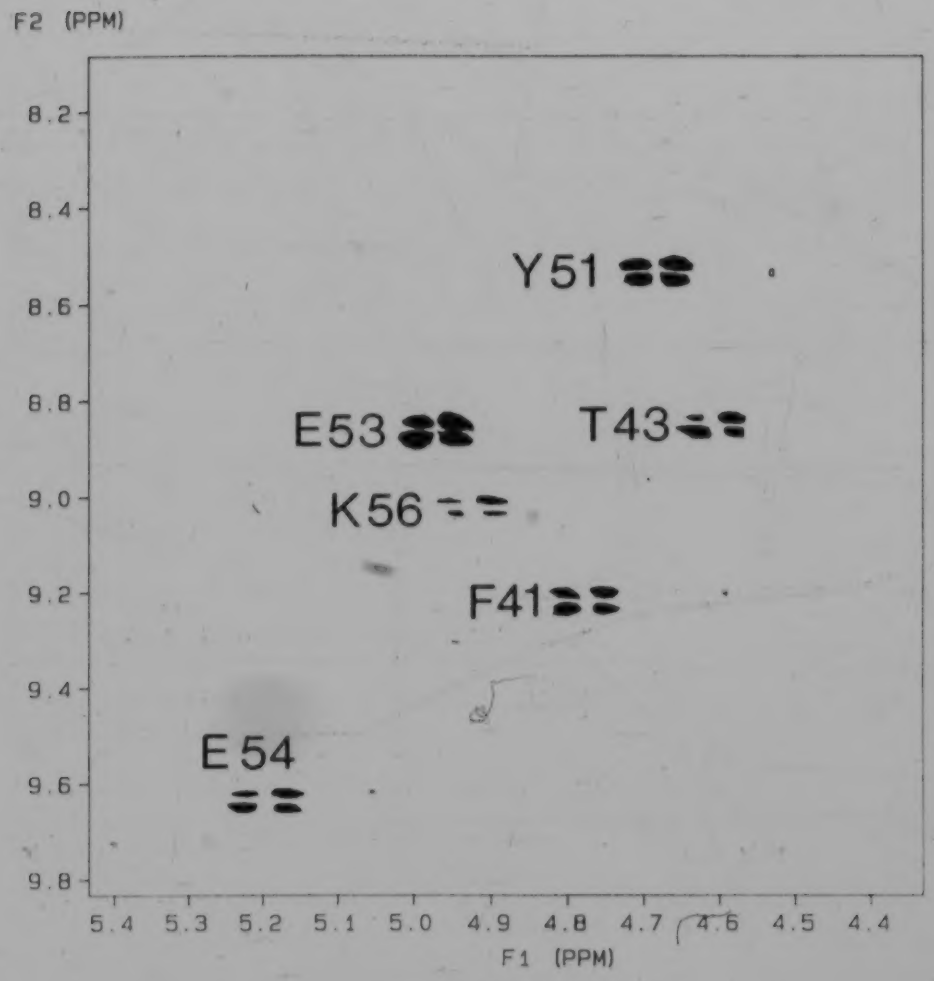


Figure V-9 DQF-COSY spectrum of V55C Cro protein

This region of the double quantum filtered COSY spectrum of V55C Cro (3 mM in standard buffer, 35°C) shows the through-bond correlations between amide protons and the α proton three bonds distant in the same residue. The cross-peaks to the six remaining amide protons after partial exchange of the protein by deuterium are indicated by residue number.

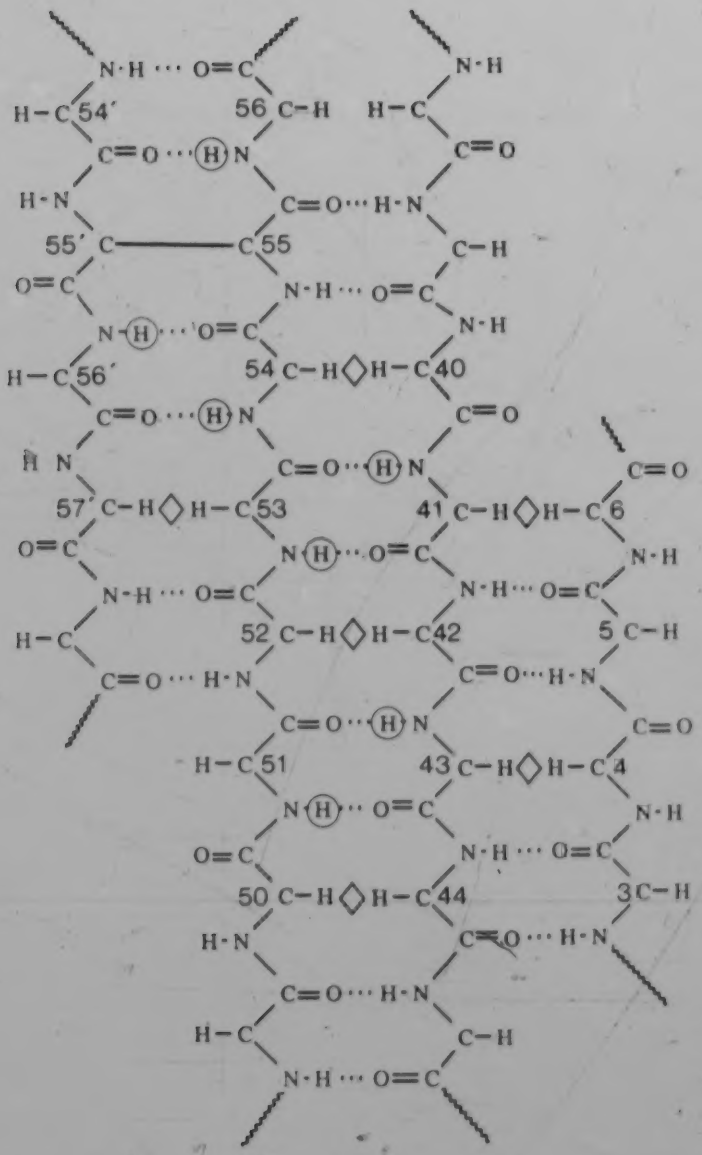


Figure V-10 Structure of the β sheet core of V55C Cro protein

A qualitative structure is shown for the central β sheet core of V55C Cro repressor protein. For clarity, R groups have not been designated and only four of the six β strands are shown. Circled amide protons are most resistant to exchange by deuterium. Diamonds indicate close approaches of α protons on adjacent strands.

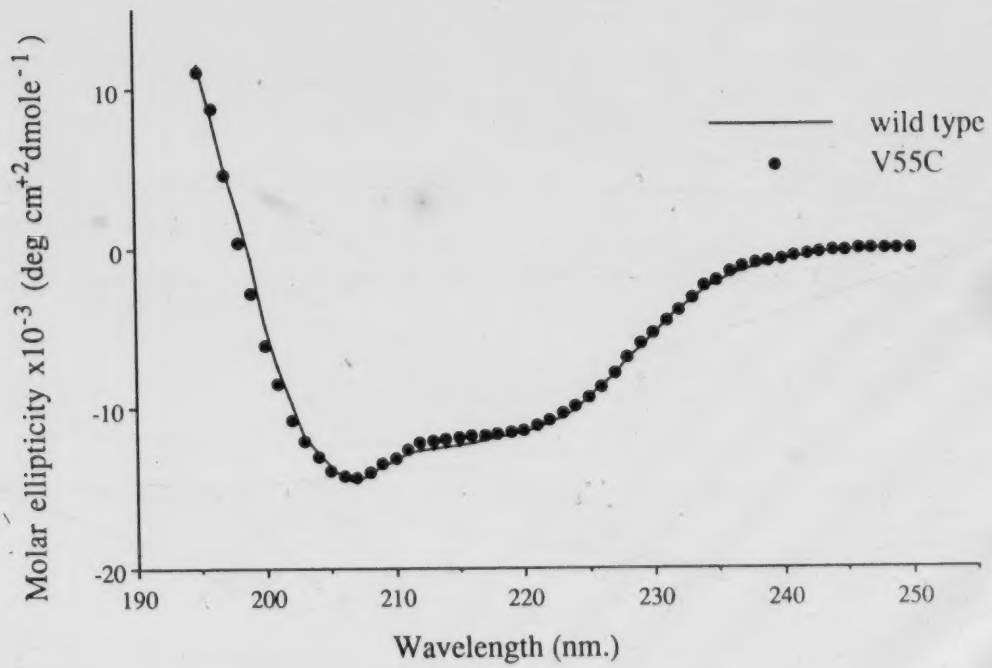


Figure V-11 CD spectra of native and V55C Cro repressor proteins

Circular dichroism spectra are shown for native Cro repressor (—), at a concentration of 27 μM, and V55C Cro (●), at a concentration of 18 μM, in standard buffer at 4.7°C.

Overall, the similar CD and ^1H NMR spectra reflect the similarity in structure. Since the two proteins have approximately the same line-widths (and the same T_2 relaxation time constants, not shown), they also have the same motional properties on the nanosecond time scale. The results below will show, however, that there are differences in flexibility between the two proteins and that these changes lead to an alteration in function.

2. Differences in structure and dynamics

a. Differences in thermal stability

The CD band intensity at 222 nm., reflecting mainly the helical content of the protein, was monitored as the temperature was allowed to rise. A drop in the magnitude of intensity reflects the loss of helical content of the two proteins. The intensity of well-resolved ^1H NMR resonances was also monitored as the temperature increased (Figure V-12). The loss of NMR intensity coincides with the loss of the interaction of chemical groups with the observed proton (mainly carbonyl groups and aromatic rings) that cause spectral dispersion. The NMR peaks reappear (in slow exchange) at the chemical shifts expected for the protein in random conformations (Bundi & Wüthrich, 1979). Results from CD and NMR[5] are summarized in Figure V-13. All melting curves were fully reversible and exhibited no hysteresis.

The increased stability for the cross-linked protein is quite apparent—the helical content of wild-type melts at 47°, V55C has its midpoint at 67°. By NMR, the wild-type protein also melts at 47°, but the V55C Cro protein melts at 58°. All NMR resonance lines give the same melting temperature to within experimental error. The cross-linked protein exists in a state between 58 and 67° with the α helices mainly intact, but without the interaction of chemical groups that cause spectral dispersion. The presence of the disulfide cross-link raises the melting temperature of Cro, presumably by de-stabilization of the unfolded state (Wetzel, 1987).

---

[5] Thermal mid-points are calculated assuming a N to D equilibrium, where N is the folded protein, and D is the denatured state. They are therefore defined as the temperature at the midpoint between the signal intensity of the folded state (after a linear correction of changes in signal intensity prior to the transition) and the signal intensity of the unfolded state. More accurately, the denaturation follows a N_2-to 2D equilibrium, and a proper description requires a protein concentration study. Such a study would also permit a better understanding of the enthalpic and entropic contributions to the denaturation process. Thermal denaturation has been monitored by NMR with approximately 2 times higher protein concentrations. The melting temperatures were the same within experimental error. Since the CD and NMR protein concentrations are not very different, the melting temperatures, as determined, can be directly compared.

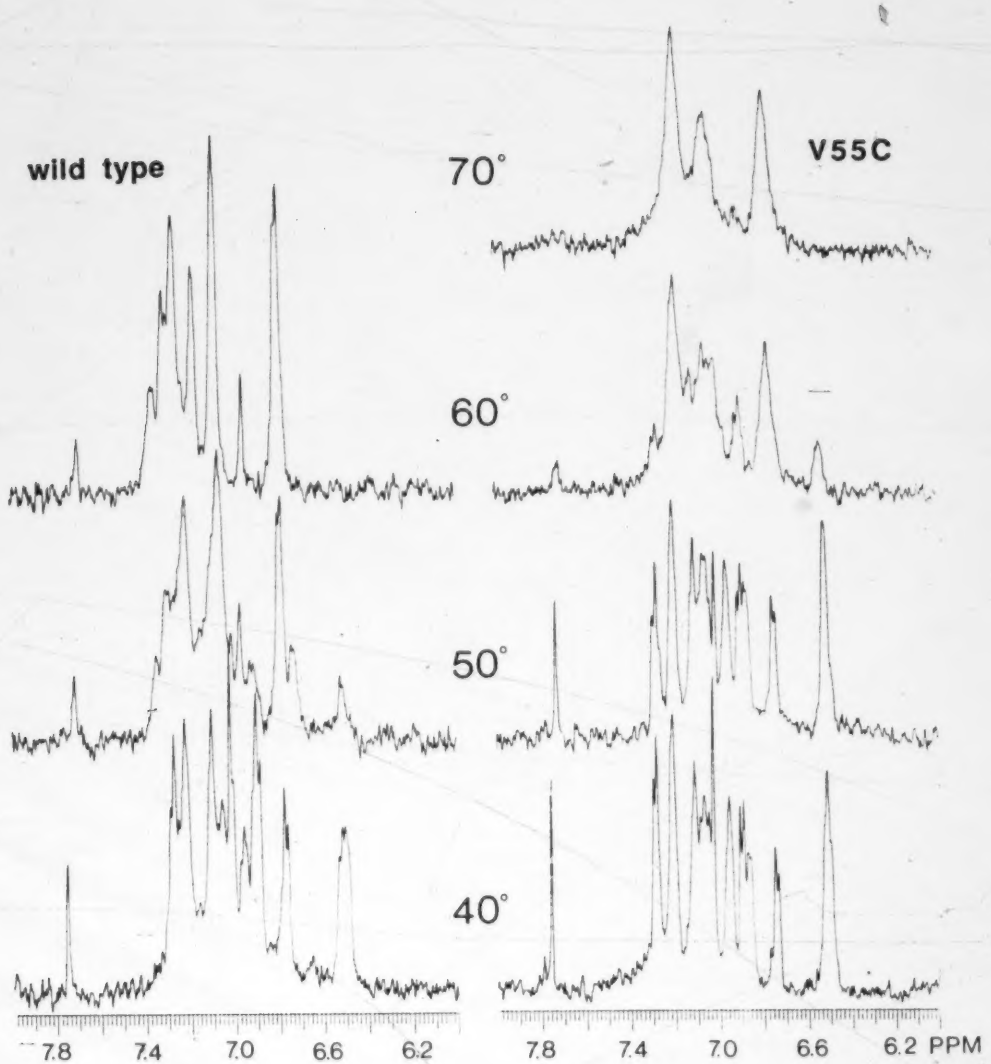


Figure V-12 Observation of thermal denaturation by NMR

Shown are ^{1}H NMR spectra of the aromatic side-chains of wild-type and V55C Cro proteins illustrating denaturation by heat. Proteins were 40 μM in standard buffer.

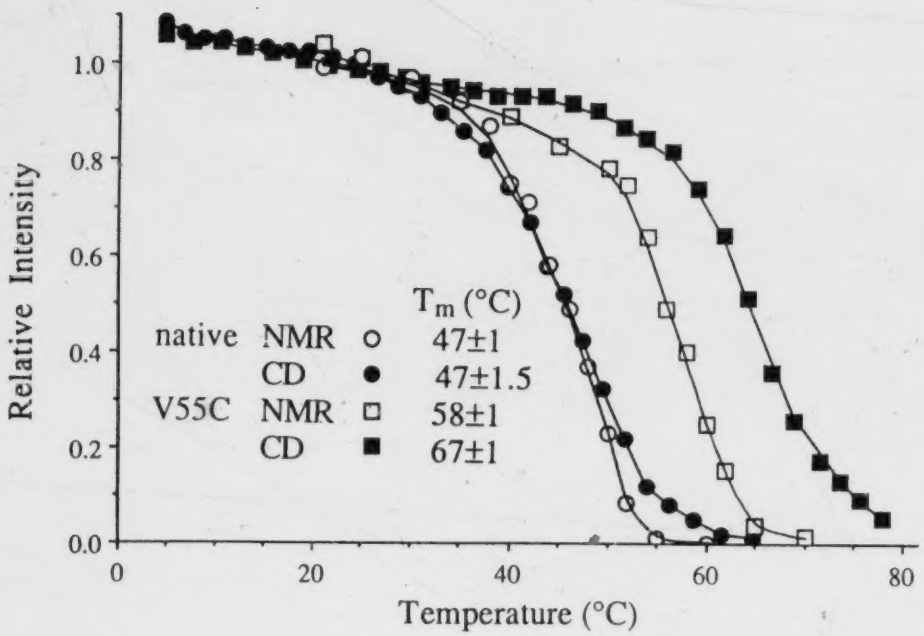


Figure V-13 Thermal stabilities of native and V55C Cro repressor proteins
Thermal denaturation profiles are shown for Cro protein. Samples were in standard buffer and proteins concentrations were: NMR/wild type, 40 μM; NMR/V55C, 40 μM; CD/wild type, 27 μM; CD/V55C, 18 μM. Thermal denaturation temperatures are defined as the midpoint between signal intensities corresponding to folded and denatured states.

b. Differences in amide proton-deuterium exchange rates

In Figure V-9, a COSY spectrum of V55C Cro shows cross-peaks between amide and α protons. The spectrum was taken approximately 24 hours after dissolving the protonated protein in D_2O at 30°C. Under these conditions, wild-type Cro loses all amide proton intensity within an hour of dissolving the lyophilized protein in D_2O. and therefore the spectrum shown in Figure V-9 would appear blank, even if spectral acquisition (about 18 hours) were initiated immediately after solvating wild-type Cro. The V55C Cro amide proton exchange rate is up to a thousand times slower than wild-type Cro (see also Chapter VII, Table VII-1), suggesting that V55C Cro cannot undergo the motion necessary for effective amide exchange.

Under conditions of low protein concentration (approximately 0.3 μM) and low ionic strength (0.05 M KCl), wild-type Cro exists as a monomer in solution. At slightly higher protein concentrations or ionic strengths the dimer form predominates (Boschelli, 1982). Even at the ionic strength and protein concentrations used in this study (0.2 M KCl, 40 μM protein) enough monomers (perhaps unfolded) of Cro may be transiently forming to allow solvent penetration and effective amide exchange. Not only is the cross-linked V55C Cro protein less flexible, it cannot have a monomer-dimer equilibrium, and is therefore resistant to amide proton exchange.

In summary, the cross-linked V55C Cro protein has qualitatively the same structure as the wild type protein. However, its increased thermal stability and resistance to amide proton exchange suggest lowered flexibility in the protein.

3. DNA-binding characteristics

a. Interaction with O_R3 operator DNA

Initial DNA-binding studies of the wild-type and cross-linked Cro repressors were carried out with the 17 base-pair O_R3 operator DNA. In Figure V-14, the CD spectra are compared for protein-DNA complexes. A slight molar excess of wild-type Cro was added to the operator DNA, producing a characteristic spectral change. The maximum difference occurs at 280 nm., which has been attributed to changes in the geometry of environment around the DNA bases (Anderson et al., 1983a), or alternatively, to an induced CD of a tyrosine residue (probably Tyr 26; Kirpichnikov et al., 1985). A nearly three fold molar excess of V55C Cro protein was added in order to generate the same magnitude in the maximum of the calculated difference spectrum, which also occurred at 280 nm.. This indicates that the wild type and V55C Cro protein interact with DNA in an similar manner.

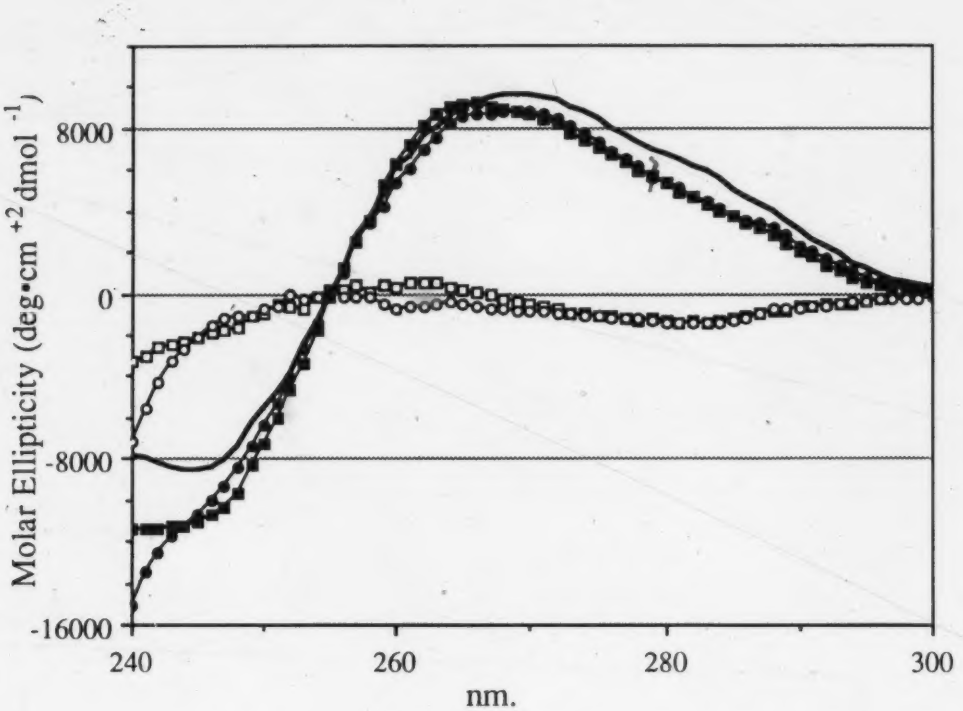


Figure V-14 CD spectra of Cro-operator DNA complexes

Circular dichroism spectra are shown for the 17 base-pair O_R3 sequences (solid line), the wild-type Cro-O_R3 complex (- ■ -), the V55C Cro-O_R3 complex (- ● -), and their respective calculated difference spectra (- □ - and - o -). The concentration of DNA used was 2 μM, wild type protein was 2.5 μM, and V55C Cro was 5.7 μM. Proteins and DNA were in standard buffer, at 25°C.

b. Interactions with O_R3 left half-operator DNA

Binding studies of the wild type and V55C Cro proteins were carried out in detail with the L10 half-operator DNA. The binding strengths and stoichiometry of protein-DNA complex formation were measured by monitoring the NMR chemical shift changes in the imino protons of the DNA (Figure V-15) as wild type or V55C Cro protein was added. The chemical shift changes of any imino proton yield the same binding constant for a given protein-DNA titration. For the binding of wild type Cro to L10 at 22°, chemical shift changes for thymine H3 are inaccurate because of the large line-widths due to exchange broadening (see below). Although titration data are shown in Figure V-16 for guanine H1 imino proton shifts, binding constants are calculated with the sum of all measurable chemical shifts in order to attain more reliable data.

Even so, the data could not be analyzed with differing δ_{max1} and δ_{max2} values (equation 6), so it was assumed that the change in signal intensity of the DNA bound was the same in both complexes ($\delta_{max1} = \delta_{max2}$). The data cannot fit 1:1 binding of the Cro dimer to L10 DNA. Subsequent titrations have confirmed 1:2 binding (see below and Chapter VI). Cooperativity effects could not be determined, although the cooperativity constant, c, was not more than 2. The binding curve fits well by assuming that the two DNA binding sites for L10 on the Cro dimer are non-interacting with c equal to 1.

The microscopic dissociation equilibrium binding constant for wild-type Cro to L10 DNA is $19\pm3 \times 10^{-6}$ M at 22°C. When concentrations of DNA and protein are within an order of magnitude of each other there are significant populations of both LP and LPL species of DNA-protein complex. As protein is added to DNA, first the LPL form is favoured; at higher protein concentrations the LP form predominates. Because the LP complex tends to be insoluble, higher protein concentrations forcing the LP complex and completely deleting populations of free DNA were not used. At concentrations of 40 μM Cro (dimer) and 40 μM DNA, (P_o/L_o = 1:1), the approximate concentrations are: free DNA, 16 μM; free Cro, 19.6 μM; Cro:L10, 16.8 μM; and Cro:$L10_2$, 3.6 μM. The relative magnitude and direction of the imino chemical shift changes at this ratio (Figure V-17) are the same as observed in the titration of the full 17 base-pair O_R3 operator by wild-type Cro (Lee et al., 1987), indicating that approximately the same changes are occurring in DNA structure when complexing with protein. The induced chemical shifts are caused by changes in the electronic environment surrounding the nucleus from hydrogen-bonding by the protein and from small alterations in DNA conformation (≤ 0.2 Å; Scheek et al., 1983).

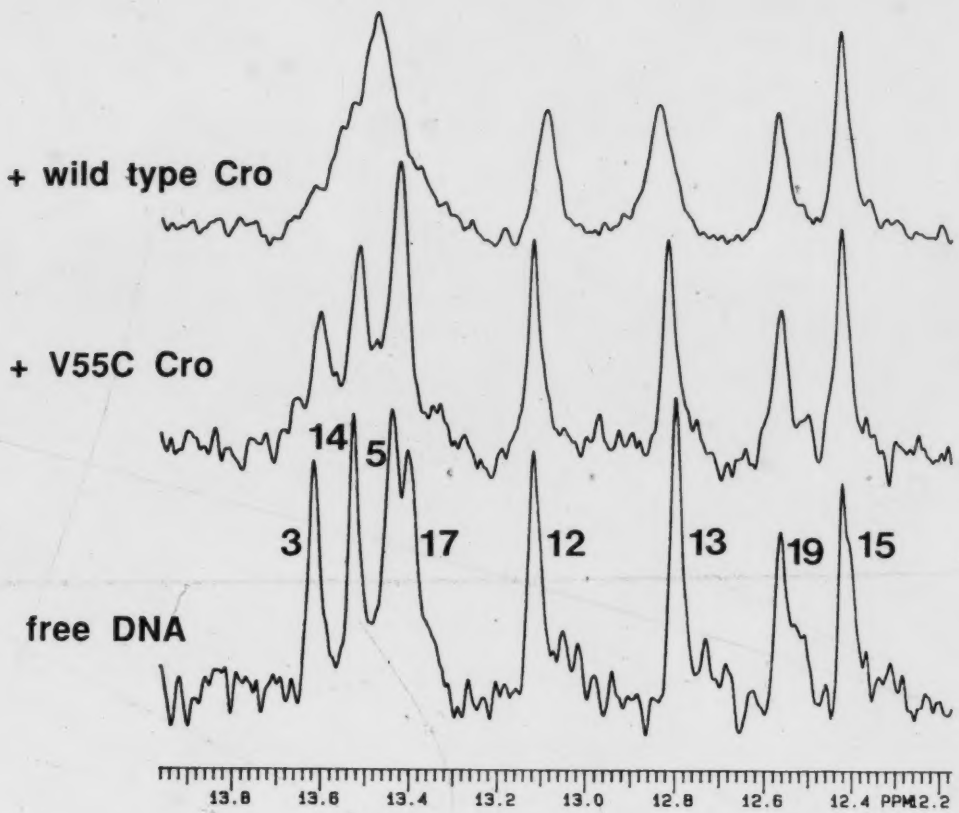


Figure V-15 Imino proton NMR resonances of free and complexed DNA
The imino proton ^1H NMR resonances of 41 μM L10 DNA in standard buffer, 85% H_2O/ 15% D_2O are shown. Numbers indicate the assignment to DNA base-pair (as defined in Chapter IV). The bottom spectrum is unbound DNA. The middle and top spectra show the imino protons when in complex with V55C and wild-type Cro proteins, respectively. The top spectrum is the sum of 8192 scans. The bottom two are 1024 scans.

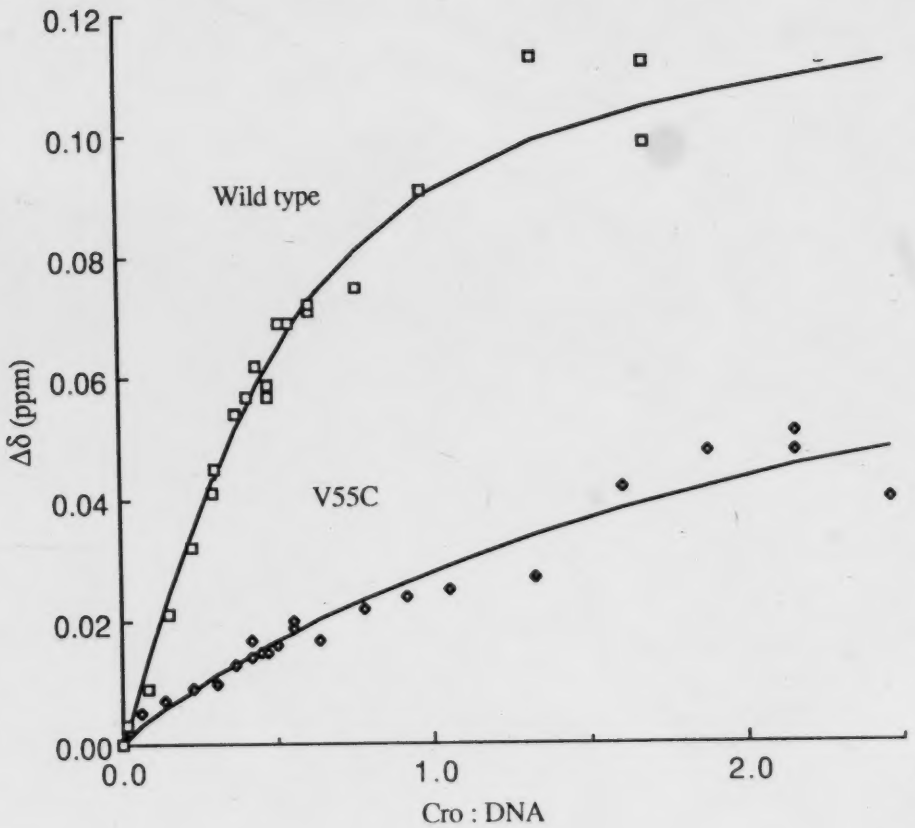


Figure V-16 Titration curves of L10 half-operator by Cro proteins
Shown are the sum of guanine imino proton chemical shift changes (nucleotides 12, 13, 15, and 19) in L10 on the addition of wild-type Cro (□) and V55C Cro (◇). The initial DNA concentration is 41 μM. Cro concentrations are for the dimeric form of the protein.

At 22°, wild-type Cro binds to L10 DNA about 8 times stronger than V55C Cro, which has a microscopic equilibrium dissociation constant of 160±30 μM. At 35°, the binding behavior of V55C is unchanged (K_d = 150 ± 50 μM). However, the binding strength of wild-type Cro substantially reduces, with K_d values changing from 19±3 to 45±10 μM. This is an indication that wild-type Cro is beginning to denature and that the cross-linked protein is functionally thermally stable. Higher temperatures could not be used, since free L10 DNA also begins to melt above 35°. With other more stable DNA sequences, it could be expected at a high enough temperature, the V55C Cro protein would bind more tightly than wild type.

At room temperature, the decrease in binding for the cross-linked protein indicates the protein may have changed conformation from the wild-type Cro even in the absence of DNA. Alternatively, the lack of flexibility in the V55C Cro may be related to an inability to make the conformational changes required in the protein monomeric unit for the interaction with one DNA half-site. In Figure V-18, the available protein signals for the wild type and V55C Cro proteins are shown both in the absence and presence of L10 DNA. In the protein-DNA complex, the molar excess of DNA over protein dimer is about 4 so that more than 90% of the protein is bound by DNA in both cases. These spectra represent 10 μM protein (dimer). Since the H_2O signal was diminished by binomial suppression, protein signals nearer to the water could not be easily seen, because of baseline distortions and non-excitation of the NMR signal. The wild-type methyl resonances change from their positions in the absence of protein. V55C Cro protein methyl resonances shift less.

Lee et al. (1987) have also shown changes in protein methyl resonances for the binding of wild type Cro to the 17 base-pair O_R3 operator DNA. They assert that these changes occur because of a conformational change in Cro required for the monomeric units to bind both halves of the operator simultaneously. The same changes occur for the binding of wild type Cro to the L10 half-operator. Therefore, these alterations in structure for residues 40 and 42 in the monomer-monomer interface region of Cro occur mainly because they are required for strong specific binding to DNA, and not only for the Cro dimer to stretch into two successive grooves of DNA.

The role of flexibility in protein-DNA interactions has been shown to be important in other DNA-protein recognition systems. The dimeric phage 434 repressor aligns the two operator DNA half-sites for binding with its two recognition helices. Although the protein does not contact the central base-pairs, the base-composition near the center of the operator affects its affinity for repressor by altering the ease with which operator DNA can be deformed into the optimal configuration for complex formation (Koudelka et al., 1988a,b). Introducing a single strand nick at the central phosphodiester bond of the operator increases the flexibility of the operator, and enhances protein-DNA binding. A more flexible mutant

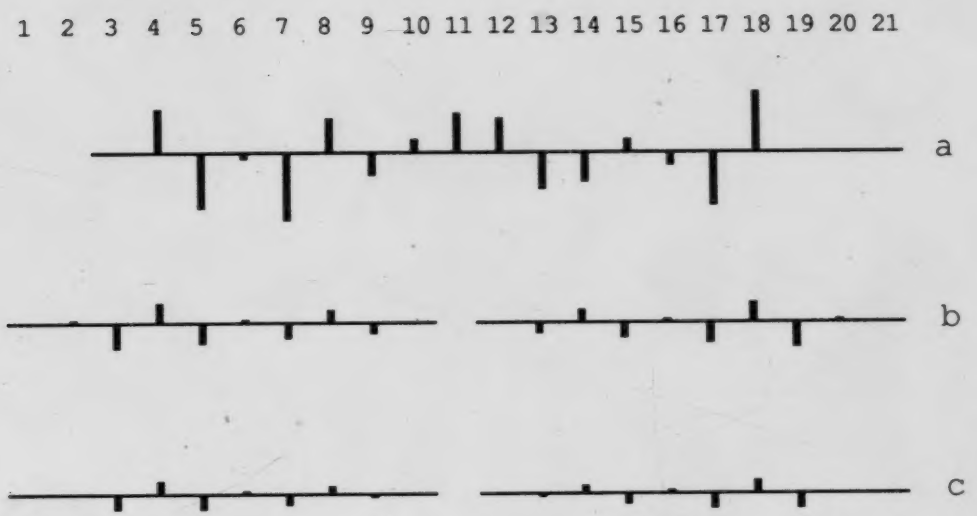


Figure V-17 Observed magnitude of imino chemical shift changes upon complex formation (a), Induced chemical shifts for the corresponding imino protons of 17 base-pair O_R3 DNA when complexed with wild-type Cro protein (adapted from Lee et al., 1987); (b), Induced chemical shifts in imino proton signals of L10 DNA when complexed to wild-type Cro protein; (c), Induced imino proton shifts when complexed with V55C Cro.

Two L10 molecules are shown. Initial DNA concentration for L10 complex formation was 41 μM in standard buffer at 22°. O_R3 complex formation had an initial DNA concentration of 200 μM in 0.3 M KCl, at 20°C. The largest arrow corresponds to a chemical shift change of 0.28 ppm. Note that complexes of L10 with wild type and V55C Cro are non-stoichiometric, and the observed chemical shift changes at a ratio of 1:1 Cro (dimer) L10 DNA are shown here. Calculated δ_{max} values are approximately 50% higher. Terminal imino protons (at positions 3/19 for the 17mer, 1/10 and 12/21 for L10) are not observed and therefore no chemical shift changes are recorded for these protons.

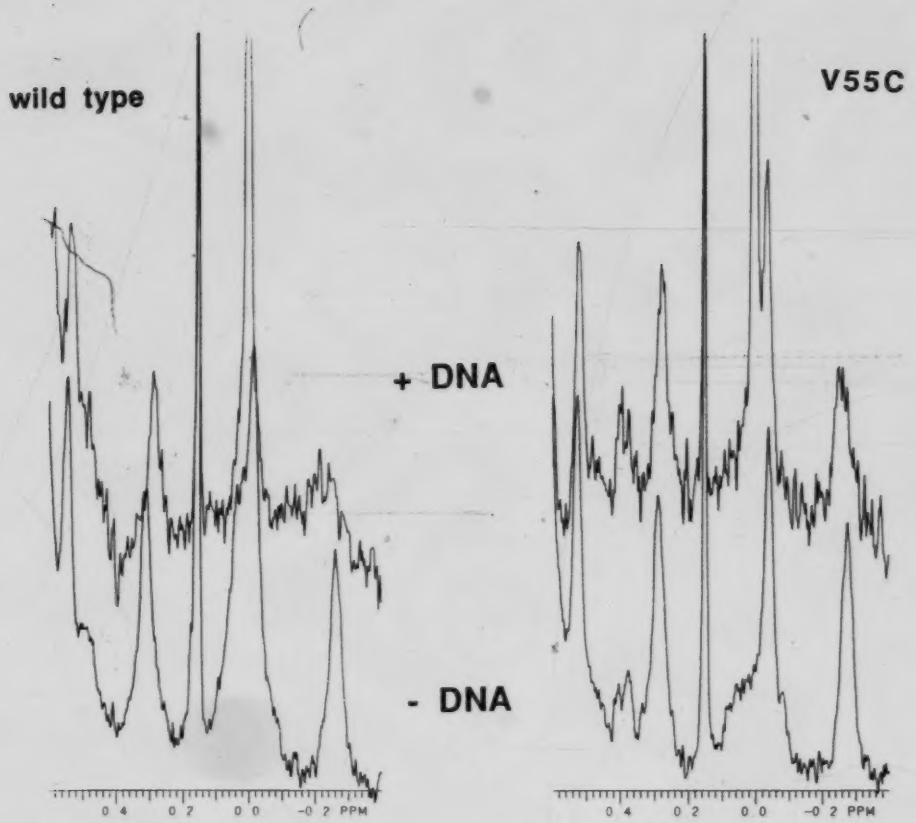


Figure V-18 Comparison of wild-type and V55C Cro proteins in protein:DNA complexes Shown are the ^{1}H NMR signals of the upfield field methyl resonances of Ile 40 and Leu 42 of wild-type Cro (left) and V55C Cro (right) in the absence (lower) and presence (upper) of a four molar excess of L10 half-operator DNA. Spectra have been apodized with 2 Hz line broadening. The protein concentrations in the lower spectra are 40 μM. One of the broad signals at 0 ppm is probably due to an impurity associated with silicone grease during lyophilization procedures. The protein concentration in protein:DNA complexes (upper spectra, MW 28,000) is 10 μM protein. All samples were in standard buffer at 22°C.

protein, having reduced interaction between protein monomers, is less sensitive to changes in flexibility of the operator.

In summary, the wild-type and V55C Cro proteins have nearly the same structure in the absence of DNA, although the cross-linked protein is more thermally stable and is resistant to backbone amide exchange. The reduction in the binding of DNA by the V55C Cro protein occurs in conjunction with its lowered flexibility.

D. Formation and structural details of the Cro-L10 DNA complex

A preliminary investigation has been made into detailed structural studies of the wild type Cro-[L10 half-operator]$_2$ protein:DNA complex. The goal of this project was to determine the structure of the L10 half-operator when bound to Cro repressor and to compare it to the structure of the free DNA (Chapter IV). After assignment of both the DNA and protein signals in the complex, it may be possible to assign inter-molecular NOEs between the protein and DNA, thereby obtaining a detailed knowledge of the interactions in the Cro protein-DNA complex.

Figure V-19 shows a region of the ^{1}H NMR spectra of L10 DNA, Cro repressor protein, and their complex. DNA aromatic H6/H8 protons resonate between 7 and 8.5 ppm., and cytosine H5 and sugar 1' protons are between approximately 5.2 ppm and 6.2 ppm.. Most of the protein resonances conveniently fall into the region between 6.2 ppm and 7 ppm, and α proton resonances are below 5.5 ppm. The spectrum of the complex shown is a molar ratio of protein (dimer) to DNA of 1:14. Some DNA resonance lines are very broad due to exchange broadening.

The titration of DNA by Cro repressor is shown in more detail in Figure V-20. At no time did any substantial amount of precipitate form during the titration, although the solution became viscous as the concentration of protein reached stoichiometric proportions. Some resonances did not appreciably broaden beyond that expected for the complex (molecular weight 28,000) and the increased viscosity of the solution. These resonances were in fast exchange, with the chemical shift observed for these line intermediate between that of free DNA and that in the protein:DNA complex. In Figure V-21, the combined shift of some of these resonances are plotted against the protein:DNA ratio. As indicated before with the titration at lower initial DNA concentrations, 2 L10 molecules can bind to the Cro dimer.

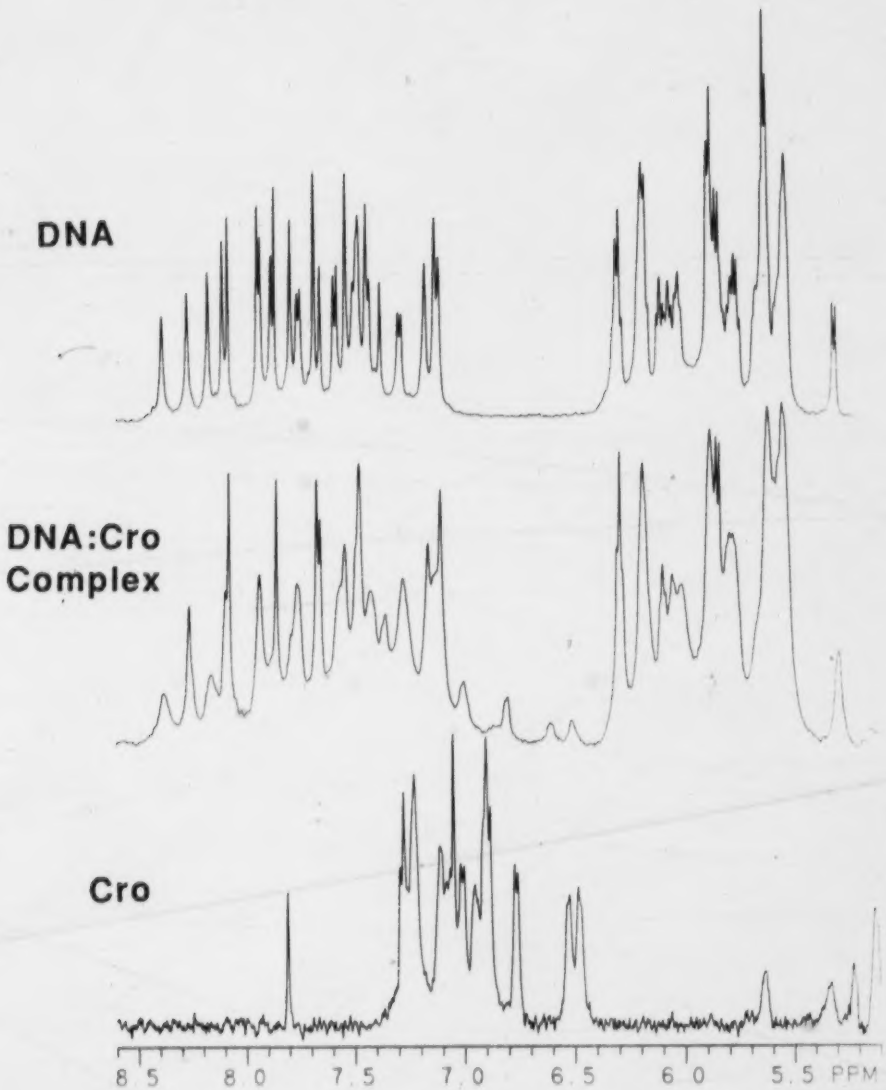


Figure V-19 ^{1}H NMR spectra of Cro, L10 DNA, and the protein:DNA complex
Shown are the ^{1}H NMR spectra between 5.2 and 8.5 ppm of Cro repressor (bottom), L10 DNA (top) and their complex (middle). The Cro protein concentration was 40 μM (dimer) in standard buffer at 20°C. DNA concentration was 2.2 mM, in 0.18 M KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 50 μM EDTA, pH 7.0, at 20°C. The complex is 2.1 mM DNA, 0.15 mM Cro protein (dimer) in the same 0.18 M KCl buffer at 20°C.

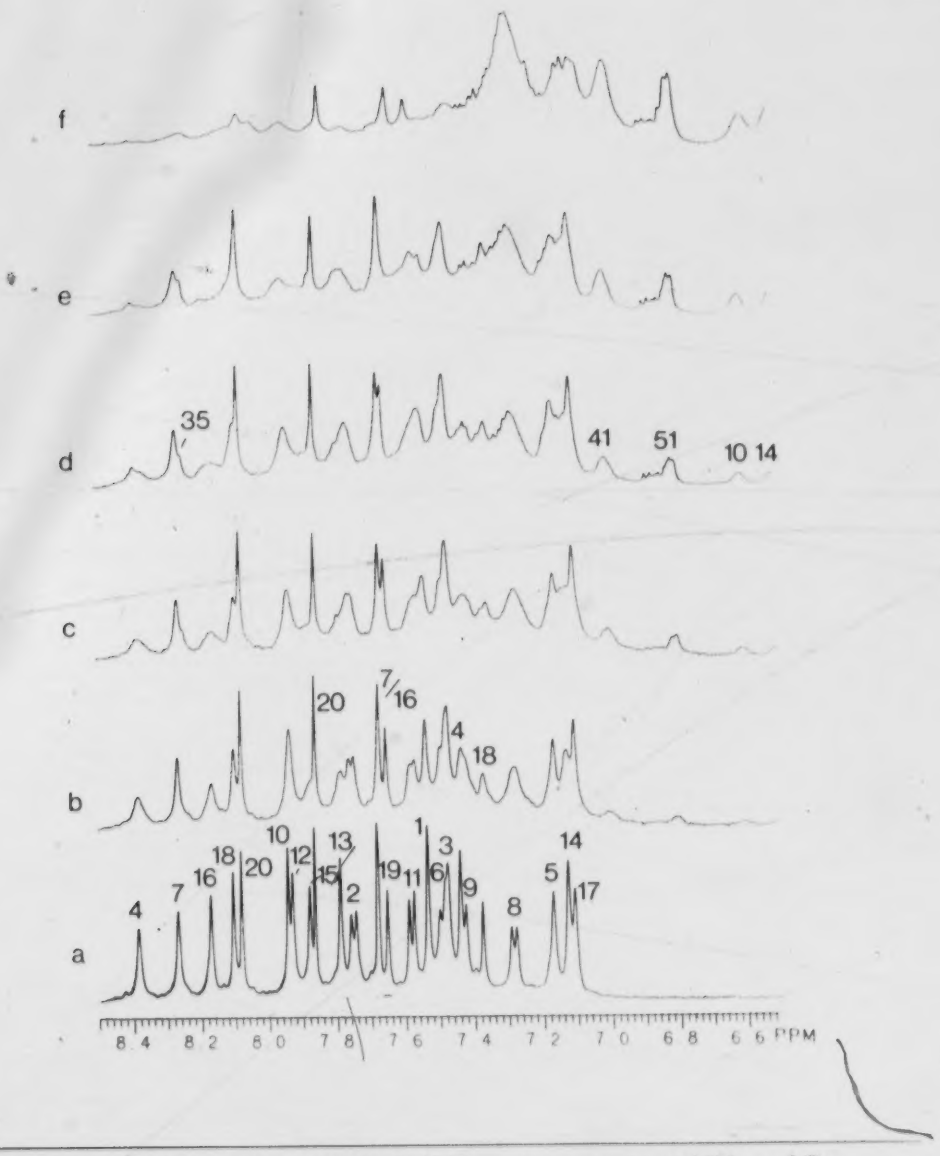


Figure V-20 Spectral changes in L10 DNA H6/H8 resonances upon addition of Cro
Shown are the ^{1}H NMR spectra between 6.5 and 8.5 ppm of DNA with increasing amounts of Cro protein added (from bottom to top). The molar ratios of protein (dimer) to L10 DNA are: (a), 0.00; (b), 0.03; (c), 0.07; (d), 0.10; (e), 0.18; (f), 0.69. The numbers indicate residue assignments for the H6 and H8 proton of the DNA on spectrum (a), the H2 protons of the DNA on spectrum (b), and the resolved proton peaks of Cro protein in spectrum (d). Buffer conditions are 0.18 M KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 50 μM EDTA, pH 7.0, at 20°C.

All of the resonances that remained in fast exchange exhibited extra broadening due to chemical exchange. Two of the resonance lines, for the A_{20} H2 and the combined A_7 and A_{16} H2 protons, were sufficiently narrow and spectrally disperse throughout the titration so that linewidths could be accurately measured and exchange broadening contributions analyzed. The calculations assumed that there are two non-interacting sites for binding of L10 on the Cro dimer. In fast exchange, the observed linewidth at half-height of the resonance, $\Delta\nu_{obs}$, is given by (Sykes & Scott, 1972):

$$2\pi\Delta\nu_{obs} = f_f(\Delta\nu_f) + f_b(\Delta\nu_b) + f_f^2 f_b^2 (\Delta\delta)^2 (\tau_f + \tau_b) \qquad (9)$$

where f_f and f_b are fractions free and bound, $\Delta\nu_f$ and $\Delta\nu_b$ refer to the linewidths of the free and bound DNA (in radians s^{-1}), $\Delta\delta$, is the chemical shift difference (in radians s^{-1}) between bound and unbound forms, and τ_f and τ_b are the lifetimes of the free and bound forms. Since the populations are related to the lifetimes by:

$$f_f = \frac{\tau_f}{\tau_f + \tau_b} \qquad \text{and} \qquad f_b = \frac{\tau_b}{\tau_f + \tau_b} \qquad (10)$$

equation 9 can be re-written as:

$$2\pi\Delta\nu_{obs} = f_f(\Delta\nu_f) + f_b(\Delta\nu_b) + f_f^2 f_b (\Delta\delta)^2 \frac{1}{k_d} \qquad (11)$$

where k_d is the dissociation rate constant ($= \tau_b^{-1}$).

The observed linewidths for the A_{20} H2 and A_7/A_{16} H2 ^{1}H NMR resonance lines are plotted against f_b in Figure V-22. Since the complex is stoichiometric at this concentration of DNA, $f_b \approx P_o/L_o$. The best fit through the experimental points gives a k_d of 1000 M^{-1} s^{-1}. Assuming an association rate constant[6] of 3×10^8 M^{-1} s^{-1}, the calculated equilibrium dissociation constant (in 0.18 M KCl) is 3.3 μM, which is consistent with the K_d measured

---

[6] The measured association rate constant for the binding of 21 base-pair O_R3 DNA to Cro is approximately 3×10^8 M^{-1} s^{-1} in 0.1 M KCl at 0°C (Kim et al., 1987). This rate is about 5 times faster than that expected from simple diffusion, presumably due to an electrostatic interaction between the positively charged protein and the negatively charged DNA. The measurements with the 10 base-pair O_R3 half-operator are at higher temperature, which would increase the rate, but at higher salt, which would tend to decrease the electrostatic interaction between Cro and DNA. The association rate constant is assumed to be approximately the same in 0.2 M KCl at 22°C.

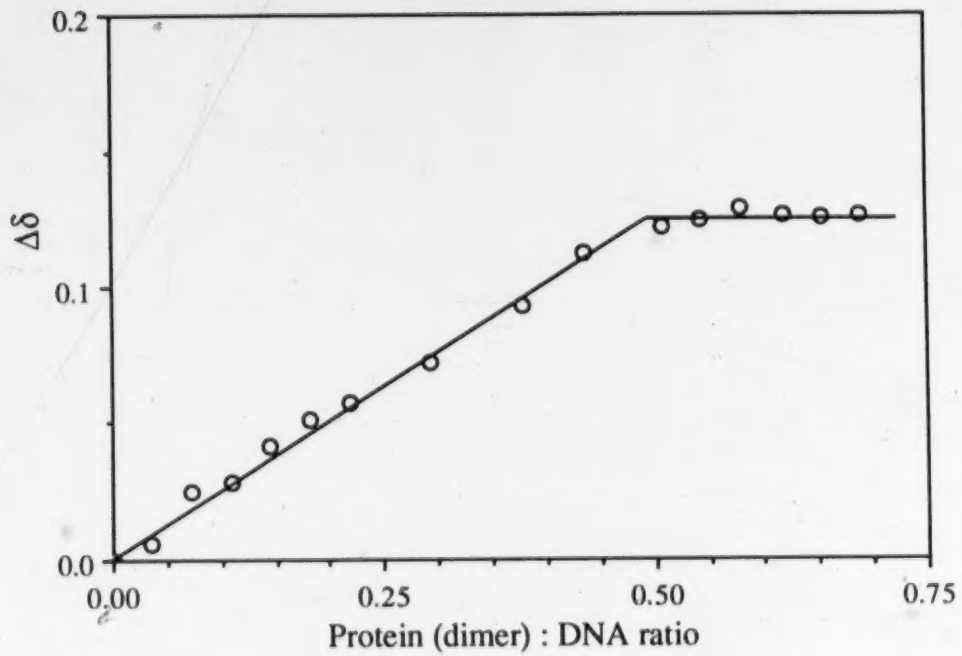


Figure V-21 Titration of L10 DNA by Cro repressor

Shown are changes in chemical shift of L10 DNA (combined shifts of A_{20} H2, A_{20} H8, $A_{7/16}$ H2 and A_7 H8) illustrating that 2 L10 molecules bind to the Cro dimer and that the complex is stoichiometric.

by following imino chemical shift changes at higher ionic strengths (19±3 μM in 0.20 M KCl). Other resonance lines cannot be analyzed in the same manner as these H2 resonances since they can not be followed throughout the titration to obtain the linewidth and chemical shift information for the protein bound complex.

Since the broadening by chemical exchange is dominating, the square root of initial increase of the linewidth of the DNA ^{1}H NMR resonances is approximately proportional to the chemical shift difference between the free and bound forms of the DNA. Several spectral lines (for A_7 H8, C_8 H5, $A_{7/16}$ H2, A_{18} H8, G_{19} H8, A_{20} H2, and A_{20} H8) are sufficiently resolved and narrow to measure the initial increase in line width (plotted as in Figure V-22) and to determine the chemical shift of the bound form. There is a reasonable correspondence between the square root of the initial increase in line width and the chemical shift difference between free and bound forms of the DNA (Figure V-23).

Initial slopes were also measured for all aromatic base H6 and H8 resonance lines plotted as in Figure V-22. The square root of this slope is *very* approximately proportional to the chemical shift difference between the free and bound forms of DNA. Although it was difficult to discern the direction of the chemical shift change for the broader lines in the spectrum presented as Figure V-20, these 'chemical shifts' are shown in Figure V-24. The chemical shifts are only approximate since they require all resonances to be in fast exchange between free and bound forms[7]. The H6 and H8 protons of the DNA were chosen since they cover the entire molecule, their ^{1}H NMR lines were reasonably well resolved, and the protons occupy roughly analogous positions in DNA, at least for unbound form (Chapter IV). In Figure V-24, there is a correspondence between the magnitude of the approximate chemical shifts and positions of proposed contacts made by Cro repressor (Ohlendorf et al., 1982). Parts of the DNA helix predicted to be not contacted by protein, such as the TCA segment of the top strand and the TAGA segment of strand 2, show little change in their H6/H8 proton chemical shifts on complexation by protein. Sequence-specific van der Waals contacts made by Cro to thymine bases appear to alter the chemical

---

[7] For example, the H8 proton of nucleotide A_4 is predicted to shift about 0.3 ppm. However, fast exchange requires $(\Delta\delta / k_d)^2 << 1$ ($\Delta\delta$ in radians s^{-1}) and 0.3 ppm = 150 Hz = 950 radian s^{-1}. Since k_d is on the order of 1000 s^{-1}, this condition is not met. Fast exchange also would predict that, for example, in spectrum (c) of Figure V-21, the A_4 H8 resonance to have shifted approximately $fb*\Delta\delta = 0.14*0.3$ ppm = 0.045 ppm. The observed shift is closer to +0.006 ppm. The line width for this spectrum is however calculated correctly by using the measured free line width of 7.4 Hz and assuming a bound line width equal to that of the observed protein lines (about 25 Hz). The square root of the initial slope is a less accurate measure for large changes in chemical shift.

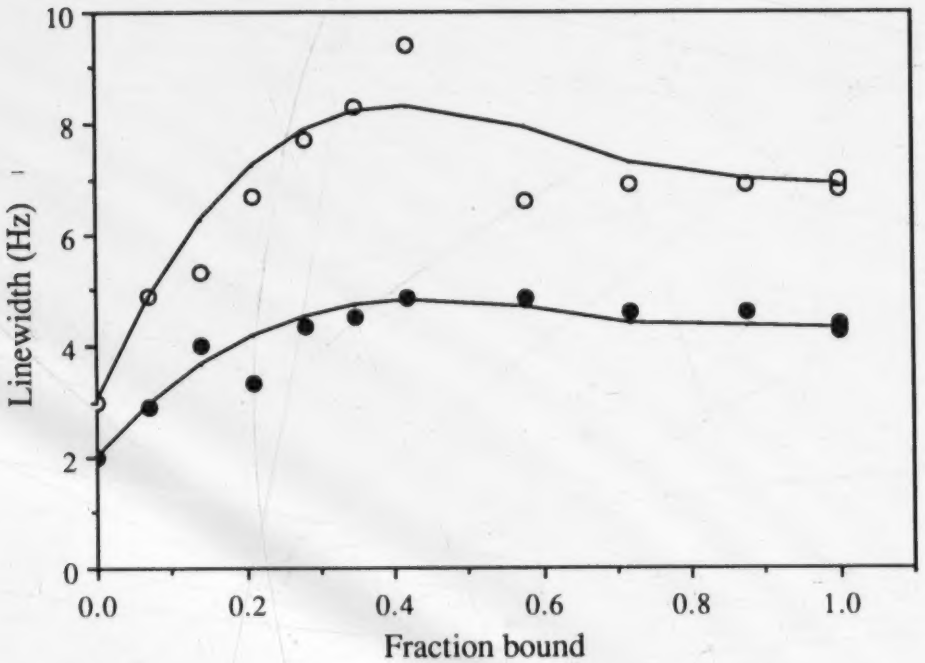


Figure V-22 Linewidth changes in DNA upon addition of Cro repressor
Shown are changes in the linewidths for the resonances of A_{20} H2 (●) and $A_{7/16}$ (O) protons of L10 DNA upon addition of Cro repressor protein. The line shows the calculated linewidth with a dissociation rate constant of 1000 s^{-1} (see text). Sample conditions are given in Figure V-21.

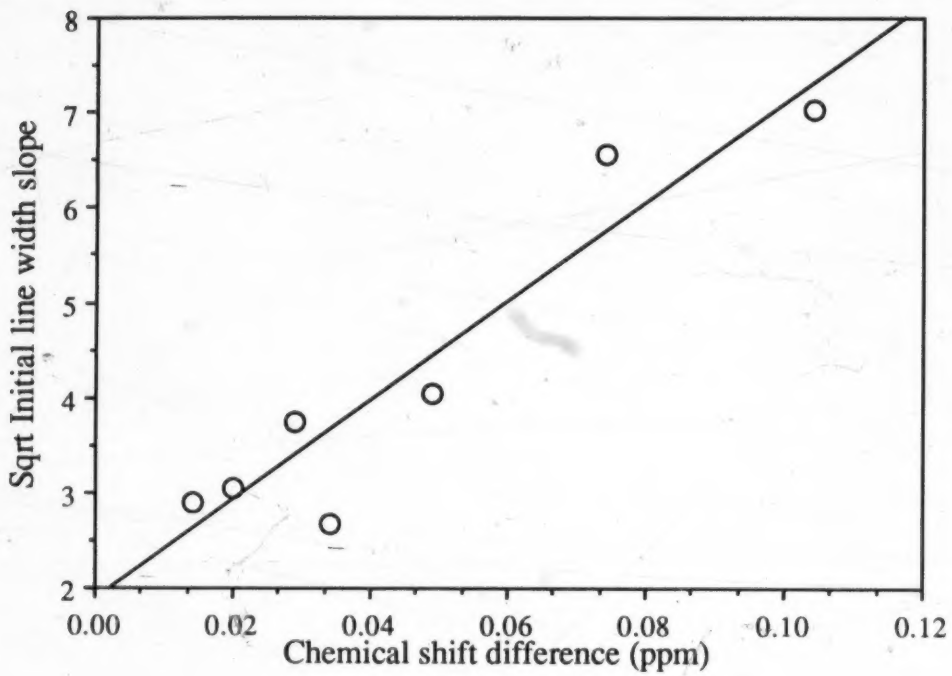


Figure V-23 Estimation of differences in chemical shift

The square root of the initial slope of the linewidth of the ^{1}H NMR resonance lines for A_7 H8, C_8 H5, $A_{7/16}$ H2, A_{18} H8, G_{19} H8, A_{20} H2, and A_{20} H8 is plotted against the observed change in chemical shift between free and bound forms of L10 DNA (see text).

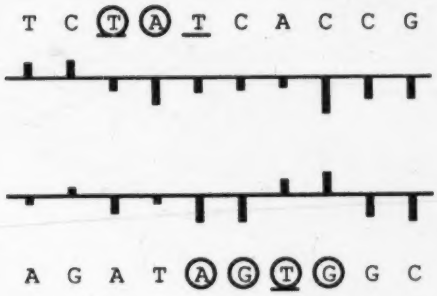


Figure V-24 DNA aromatic base chemical shift changes upon addition of Cro repressor
Shown are approximate changes in chemical shift of the H6 and H8 protons of L10 DNA between the unbound form and when in complex with Cro repressor. The direction of the chemical shift change is uncertain for shifts of large magnitude. Bases proposed in the model by Ohlendorf et al. (1982) to be hydrogen-bonded by the protein are circled. Van der Waals contacts to thymine methyl groups are underlined. Proposed non-specific contacts between protein and DNA have not been included.

shift less than hydrogen bonds. Structural changes nearer the center of the operator (base-pairs 9 and 10 of L10 DNA) are reflected in chemical shift changes here, despite the lack of direct protein contact. Unlike the full operator containing two half-sites, L10 DNA is not required to bend to fit the Cro dimer for alignment of the two half-sites of the DNA with the two recognition helices of Cro. This indicates that the structural changes seen in the crystal structure of the Cro:O_R3 complex (Brennan & Matthews, 1989b) do not completely arise from the requirement for aligning two half-sites of DNA for binding to the dimer but at least partially from changes induced by the hydrogen bonds and van der Waals contacts made by Cro at the consensus binding site (the circled DNA residues of Figure V-24).

From the linewidths of the stoichiometric complex of 2 L10 and Cro (Figure V-20) it is obvious that full resonance assignment and structural determination will be very difficult since NOESY and especially COSY experiments are even more sensitive to losses in signal intensity than one-dimensional spectra (Weiss et al., 1984). Even structure determination of the DNA with a small amount of Cro (e.g., a 1:14 Cro dimer:DNA ratio) will not be easy, as spectral lines are affected differently by exchange broadening contributions. NOESY spectra corresponding to the one-dimensional spectra in Figure V-20 are shown in Figure

IV-2 (of Chapter IV) for free DNA, Figure V-25 for unbound Cro repressor, and Figure V-26 for the DNA:protein complex.

In Figure V-26, some NOESY cross-peaks are visible showing the close contacts made by protons within the bound protein. These include interaction of the Phe 14 and Tyr 10 rings, and all three Tyr *ortho-meta* ring cross-peaks. In comparison to the NOESY spectrum of the unbound protein, it is easily seen that from the lack of changes in NOE cross-peaks and chemical shifts the protein retains at least part of its unbound structure in the complex with DNA. A notable exception is Tyr 26, whose *meta* protons shift downfield about 0.25 ppm and *ortho* protons about 0.35 ppm. The resonance position shifts for Tyr 26 (and His 35) are approximately the same in the L10-Cro complex as in the 17 base-pair O_R3 DNA:Cro complex (Lee et al., 1987) and a different half-operator (L9):Cro complex (see also Chapter VI). The Tyr *ortho-meta* cross-peaks are confirmed in the COSY spectrum (not shown).

NOESY spectra with the 1:14 Cro (dimer):DNA ratio offers a number of advantages in studying protein-nucleic acid interactions using NMR techniques. At this protein concentration, the solution is not much more viscous than that of free Cro protein, unlike the situation with the 1:2 Cro:DNA complex. Also, since the fraction of bound DNA is 2*1/14, (about 0.14), the resonance positions of the DNA are nearly the same as the uncomplexed DNA. Unfortunately, no inter-molecular NOE cross-peaks can yet be distinguished between Cro and L10 DNA, which would confirm the Ohlendorf et al. (1982) model, and be complementary to the X-ray structural work in progress. For example, the peak at 7.07 and 8.12 ppm (in the F1 and F2 PPM dimensions, respectively), is likely between the H8 proton of A_{18} and the H6 proton of T_{17} and is not between protein and DNA. Without doubt, parts of the DNA contacted by the protein are altered in structure and large chemical shift changes occur. Under non-stoichiometric conditions, exchange broadening produces excessive linewidths and cross peaks would be of exceptionally low signal to noise in the NOESY spectrum. The stoichiometric complex cannot be easily studied since at higher protein:DNA ratios significant populations of the single DNA:Cro dimer complex forms, leading to sample aggregation and broad resonance lines. For example, in Figure V-20, the His 35 proton on the C2 carbon of the aromatic side-chain becomes very broad at higher protein:DNA ratios (see also Chapter VI). The key to a complete structure determination of a Cro:DNA complex using NMR techniques will be to change the equilibrium dissociation constant by an order of magnitude to force the complex to either show resonance lines in fast exchange or in slow exchange, and to design the system so that aggregation does not occur. Heteronuclear techniques, such as ^{19}F incorporation into the DNA and ^{15}N isotope incorporation into protein simplify the NMR

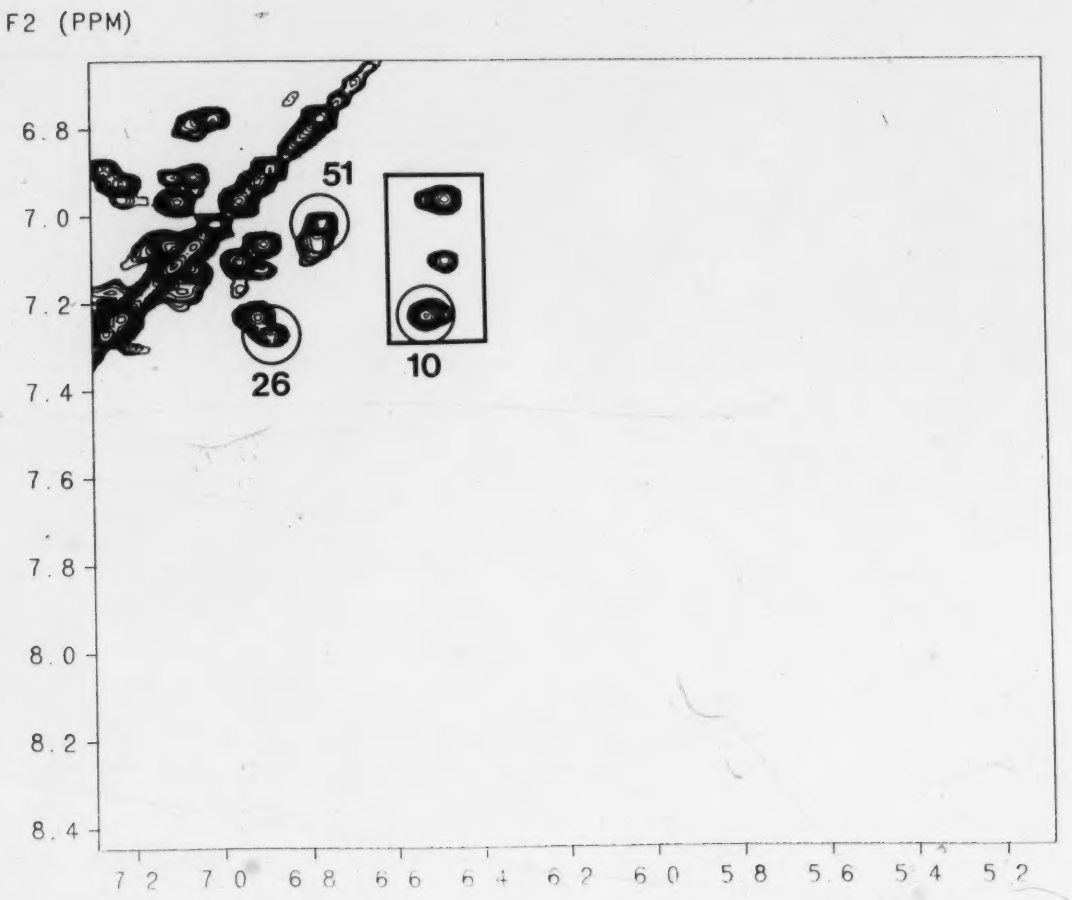


Figure V-25 NOESY spectrum of unbound Cro repressor

The region of the NOESY spectrum shown includes the aromatic side-chain resonances of Cro repressor and includes an empty area where DNA H6/H8 base-sugar 1' protons NOESY cross-peaks would occur (see Figure V-26). The protein was 0.75 mM in 50 mM KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 75 μM EDTA, pH 6.9. Tyr *ortho-meta* cross-peaks are circled and cross-peaks illustrating the interaction within and between the aromatic side-chains of Tyr 10 and Phe 14 are boxed.

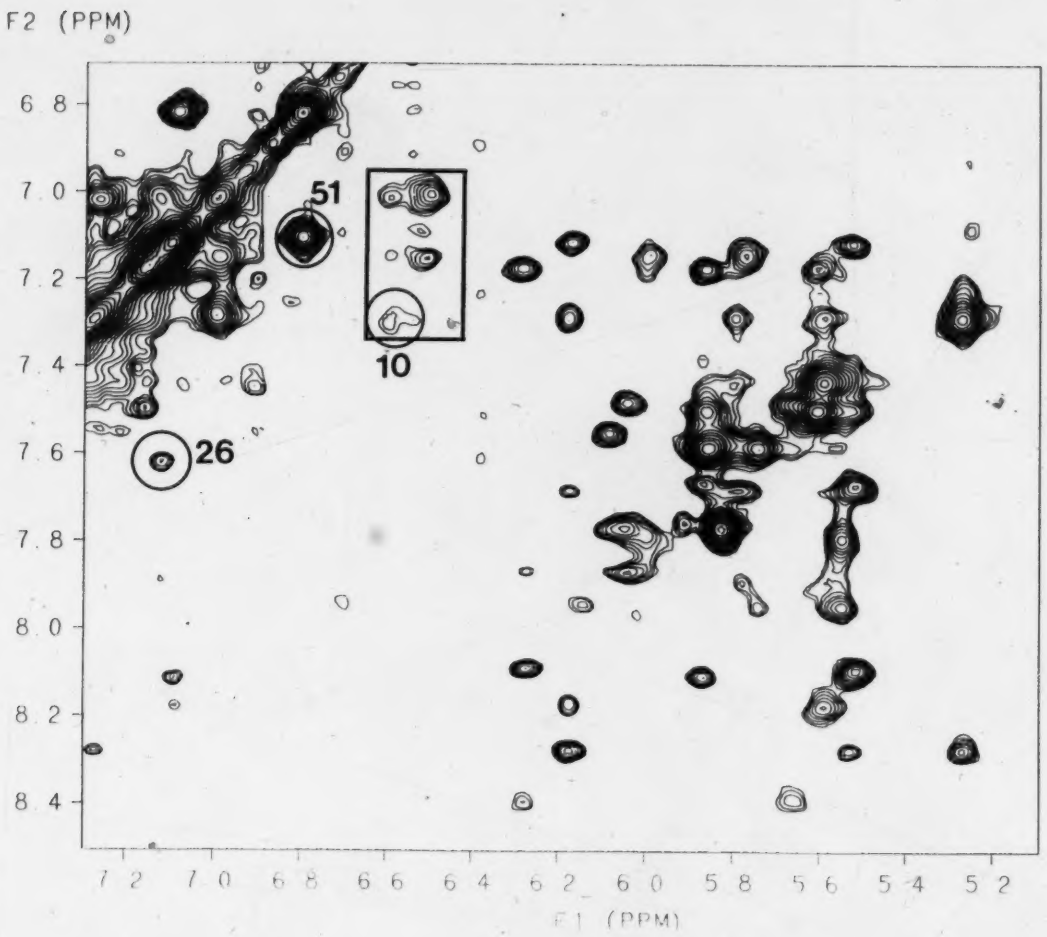


Figure V-26 NOESY spectrum of the Cro:DNA (1:14) complex

Shown region of the NOESY spectrum showing the aromatic side-chains of Cro protein, and the aromatic base and sugar 1' protons of L10 DNA. Protein was 0.15 mM (dimer) and L10 DNA was 2.1 mM in 0.18 KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 50 μM EDTA, pH 7.0 at 20°C. Assignment of the DNA aromatic and 1' sugar protons is illustrated in Figure IV-2 of Chapter IV. Tyr *ortho-meta* cross-peaks are circled and cross-peaks illustrating the interaction within and between the aromatic side-chains of Tyr 10 and Phe 14 are boxed.

spectrum, and have been used to study Cro-DNA interaction (Metzler & Lu, 1989; Leighton & Lu, 1987). The most detailed study of a protein-nucleic acid system to date using NMR techniques has been with the Lac repressor headpiece for which the DNA-binding domain has sufficiently weak DNA affinity so that NMR lines remain in fast exchange and components of the system are sufficiently soluble so that high concentrations can be used without appreciable aggregation (Lamerichs et al., 1989).

In summary, the specific complex between DNA and Cro repressor protein has been studied by ^{1}H NMR and CD spectroscopies. The structures of wild-type phage λ Cro and a mutant V55C Cro were found to be similar in the absence of DNA and at room temperature. However, the cross-linked V55C Cro is less flexible than wild-type and binds L10 DNA less tightly. L10 DNA is shown to bind to wild-type Cro in a ratio of 2 DNA duplexes per Cro dimer. Linewidth changes induced by exchange broadening indicate a dissociation rate constant of 1000 s^{-1}, consistent with an equilibrium dissociation constant of 3.3 μM and an association rate constant of about 3×10^8 M^{-1} s^{-1}. ^{1}H NMR spectral line-width broadening in the DNA approximately corresponds to the base-pairs proposed by the Ohlendorf et al. (1982) model to be contacted by the Cro repressor protein. Techniques needed for a detailed structure determination of the DNA within the Cro-DNA complex were indicated and include avoidance of intermediate chemical shift exchange regimes and aggregation of the protein-DNA complex.

E. References

Anderson, W. F., Ohlendorf, D. H., Cygler, M., Takeda, Y., & Matthews, B. W. (1983a) in *Gene Expression*, pp 19-32, Liss, New York.

Anderson, W. F., Cygler, M., Vandonselaar, M., Ohlendorf, D. H., Matthews, B. W., Kim, J., & Takeda, Y. (1983b) *J. Mol. Biol. 168*, 903-906.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754-758.

Anderson, W. F., Takeda, Y., Echols, H., & Matthews, B. W. (1979) *J. Mol. Biol. 130*, 507-510.

Anderson, W. F., Takeda, Y., Ohlendorf, D. H., & Matthews, B. W. (1982) *J. Mol. Biol. 159*, 745-751.

Arndt, K. T., Boschelli, F., Cook, J., Takeda, Y., Tecza, E., & Lu, P. (1983) *J. Biol. Chem. 258*, 4177-4183.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504-1509.

Benson, N., & Youderian, P. (1989) *Genetics 121*, 5-12.

Berg, O. G., & von Hippel, P. H. (1988) *Trends Biochem. Sci. 13*, 207-211.

Bolotina, I. A., Kurochkin, A. V., & Kirpichnikov, M. P. (1983) *FEBS Letters 155*, 291-294.

Boschelli, F. (1982) *J. Mol. Biol. 162*, 267-282.

Boschelli, F., Arndt, K., Nick, H., Zhang, Q., Lu, P., & Takeda, Y. (1982) *J. Mol. Biol. 162*, 251-266.

Brennan, R. G., & Matthews, B. W. (1989a) *J. Biol. Chem. 264*, 1903-1906.

Brennan, R. G., & Matthews, B. W. (1989b) *Trends Biochem. Sci. 14*, 286-290.

Brennan, R. G., Takeda, Y., Kim, J., Anderson, W. F., & Matthews, B. W. (1986) *J. Mol. Biol. 188*, 115-118.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285-297.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*, pp 435-428.

Caruthers, M. H., Barone, A. D., Beltman, J., Bracco, L. P., Dodds, D. R., Dubendorff, J. W., Eisenbeis, S. J., Gayle, R. B., Prosser, K., Rosendahl, M. S., Sutton, J., & Tang, J.-Y. (1986) *in Protein Structure, Folding, and Design* (Oxender, D. L., ed.) pp. 221-228, Liss, New York.

Caruthers, M. H., Gottlieb, P., Bracco, L., & Cummins, L. (1987) in *Protein Structure, Folding, and Design 2* (Oxender, D. L., ed.) pp. 9-24, Liss, New York.

Chou, S.-H., Hare, D. R., Wemmer, D. E., & Reid, B. R. (1983) *Biochemistry 22*, 3037-3041.

Cygler, M., & Anderson, W. F. (1986) *J. Biol. Struct. Dynam. 3*, 1055-1065.

Dodd, I. B., & Egan, J. B. (1977) *J. Mol. Biol. 194*, 557-564.

Eisenbeis, S. J., Nasoff, M. S., Noble, S. A., Bracco, L. P., Dodds, D. R., & Caruthers, M. H. (1985) *Proc. Natl. Acad. Sci. 82*, 1084-1088.

Englander, S. W., & Kallenbach, N. R. (1984) *Quat. Rev. Biophys. 16*, 521-555.

Ferguson-Miller, S. & Koppenol, W. H. (1981) *Trends Biochem. Sci. 6*, pp IV-VII.

Folkmanis, A., Takeda, Y., Simuth, J., Gussin, G., & Echols, H. (1976) *Proc. Natl. Acad. Sci. 73*, 2249-2253.

Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Biophys. J. 16*, 279-286.

Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Skryabin, K. G., & Kirpichnikov, M. D. (1985) *Eur. Biophys. J. 12*, 87-95.

Hayes, J. J., & Tullius, T. B. (1989) *Biochemistry 28*, 9521-9527.

Hochschild, A., & Ptashne, M. (1986) *Cell 44*, 925-933.

Hochschild, A., Douhan, J., & Ptashne, M. (1986) *Cell 47*, 807-816.

Hsiang, M. W., Cole, R. D., Takeda, Y., & Echols, H. (1977) *Nature 270*, 275-277.

Hull, W. E. (1975) Ph. D. Thesis, Harvard University, Cambridge, U.S.A..

Iwahashi, H., Akutsu, H., Kobayashi, Y., Kyogoku, Y., Ono, T., Koga, H., & Horiuchi, T. (1982) *J. Biochem. (Tokyo) 91*, 1213-1221.

Johnson, A., Meyer, B. J., & Ptashne, M. (1978) *Proc. Natl. Acad. Sci. 75*, 1783-1787.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149-158.

Kirpichnikov, M. P., Kurochkin, A. V., & Skryabin, K. G. (1982a) *FEBS Letters 150*, 407-410.

Kirpichnikov, M. P., Kurochkin, A. V., & Bayev, A. A. (1982b) *Dokl. Akad. Nauk SSSR, 266*, #3, 738-741; Engl. transl. 255-258.

Kirpichnikov, M. P., Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Kurochkin, A. V., Skryabin, K. G., & Bayev, A. A. (1984a) *Nucl. Acids Res. 12*, 3551-3561.

Kirpichnikov, M. P., Kurochkin, A. V., Chernov, B. K., & Skryabin, K. G. (1984b) *FEBS Letters 175*, 317-320.

Kirpichnikov, M. P., Yartzev, A. P., Minchenkova, L. E., Chernov, B. K., & Ivanov, V. I. (1985) *J. Biol. Struct. Dynam. 3*, 529-536.

Koudelka, G. B., Harbury, P., Harrison, S. C., & Ptashne, M. (1988a) *Proc. Natl. Acad. Sci. 85*, 4633-4637.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1988b) *Nature 326*, 886-888.

Kurochkin, A. V. & Kirpichnikov, M. P. (1982) *FEBS Letters 150*, 411-415.

Kurochkin, A. V., & Kirpichnikov, M. P. (1986) *Molekul. Biologiya 20*, #4, 985-993; Engl. transl. 798-806.

Kurochkin, A. V., Bushvev, U. N., Sepetov, N. F., & Kirpichnikov, M. P. (1986) *Molekul. Biologiya 20*, #4, 974-984; Engl. transl. 788-797.

Laemmli, U. K. (1970) *Nature 227*, 680-685.

Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Rüterjans, H. (1989) *Biochemistry 28*, 2985-2991.

Lee, S. J., Akutsu, H., Kyogoku, Y., Kitano, K., Tozuka, Z., Ohta, A., Ohtsuka, E., Ikehara, M. (1983) *Nucl. Acid Res. Symp. Series 12*, 197-200.

Lee, S. J., Shirakawa, M., Akutsu, H., Kyogoku, Y., Shiraishi, M., Kitano, K., Shin, M., Ohtsuka, E., & Ikehara, M. (1987) *EMBO J. 6*, 1129-1135.

Leighton, P., & Lu, P. (1987) *Biochemistry 26*, 7262-7271.

Maniatis, T., Fritsch, E. F., & Sanbrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., & Takeda, Y. (1982) *Proc. Natl. Acad. Sci. 79*, 1428-1432.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., & Takeda, Y. (1983) *Trends Biochem. Sci. 8*, 25-29.

Mayer, R., Lancelot, G., & Hélène, C. (1983) *FEBS Letters 153*, 339-344.

Metzler, W. J., & Lu, P. (1989) *J. Mol. Biol. 205*, 149-164.

Metzler, W. J., Arndt, K., Tecza, E., Wasilewski, J., & Lu, P. (1985) *Biochemistry 24*, 1418-1424.

Ohlendorf, D. H., & Matthews, B. W. (1983) *Ann. Rev. Biophys. Bioeng. 12*, 259-284.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718-723.

Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O., & Matthews, B. W. (1983a) *J. Mol. Biol. 169*, 757-769.

Ohlendorf, D. H., Anderson, W. F., Takeda, Y., & Matthews, B. W. (1983b) *J. Bio. Struct. Dynam. 1*, 553-563.

Oikawa, K., Kay, C. M., & McCubbin, W. D. (1968) *Biochim. Biophys. Acta 168*, 164-167.

Pabo, C. O., & Sauer, R. T. (1984) *Ann. Rev. Biochem. 53*, 293-321.

Pakula, A. A., & Sauer, R. T. (1989) *Proteins 5*, 202-210.

Pakula, A. A., Young, V. B., & Sauer, R. T. (1986) *Proc. Natl. Acad. Sci. 83*, 8829-8833.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M., & Sauer, R. T. (1980) *Cell 19*, 1-11.

Roberts, T. M., Kacich, R., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci 76*, 760-764.

Roberts, T. M., Shimatake, H., Brady, C., & Rosenberg, M. (1977) *Nature 270*, 274-275.

Sarai, A., & Takeda, Y. (1987) in *Proteins, Structure, Folding and Dynamics 2*, pp 57-64, Liss, New York.

Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom, J. H., Kaptein, R., Rüterjans, H., & Bayreuther, K. (1983) *Biochemistry 22*, 228-235.

Shea, M. A., & Ackers, G. K. (1985) *J. Mol. Biol. 181*, 211-230.

Shirakawa, M., Kawata, Y., Lee, S. J., Akutsu, H., Sakiyama, F., & Kyogoku, Y. (1985a) *J. Biochem. (Tokyo) 98*, 799-805.

Shirakawa, M., Lee, S. J., Akutsu, H., Kyogoku, Y., Kitano, K., Shin, M., Ohtsuka, E., & Ikehara, M. (1985b) *FEBS Letters 181*, 286-290.

Sykes, B. D., & Scott, M. D. (1972) *Ann. Rev. Biophys. Bioeng. 1*, 27-50.

Takeda, Y. (1979) *J. Mol. Biol. 127*, 177-187.

Takeda, Y., Folkmanis, A., & Echols, H. (1977) *J. Biol. Chem. 252*, 6177-6183.

Takeda, Y., Kim, J. G., Caday, C. G., Steers, E., Jr., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1986) *J. Biol. Chem. 261*, 8608-8616.

Takeda, Y., Ohlendorf, D. H., Anderson, W. F., Matthews, B. W. (1983) *Science 221*, 1020-1026.

Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci 86*, 439-443.

Tullius, T. D., & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. 83*, 5469-5473.

Ulrich, E. L., John, E.-M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., & Markley, J. L. (1983) *Biochemistry 22*, 4362-4365.

Weber, P. L., Drobny, G., & Reid, B. R. (1985a) *Biochemistry 24*, 4549-4552.

Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985b) *Biochemistry 24*, 4553-4562.

Weiss, M. A., Eliason, J. L., & States, D. J. (1984) *Proc. Natl. Acad. Sci. 81*, 6019-6023.

Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Proc. Natl. Acad. Sci. 81*, 130-134.

Wemmer, D. E., Shou, S.-H., & Reid, B. R. (1984) *J. Mol. Biol. 180*, 41-60.

Wetzel, R. (1987) *Trends Biochem. Sci. 12*, 478-482.

Williams, T. C., Corson, D. C., Oikawa, K., McCubbin, W. D., Kay, C. M., & Sykes, B. D. (1986) *Biochemistry 25*, 1835-1846.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic acids*, Wiley, New York.

Chapter VI

Asymmetry in the Structure and Function of Phage λ O_R3 Operator DNA

A. Introduction

The Cro repressor, like many regulators of transcription, is a dimeric protein composed of identical subunits that recognizes and binds to a specific site on DNA (Schlief, 1988; Pabo & Sauer, 1984). The dimeric protein binding site, the operator, has two-fold symmetry in its sequence, and can be thought of being composed from two nearly identical half-operators (cf. Figure I-4). The X-ray crystallographic determination of the Cro repressor revealed two symmetry-related α-helices protruding from the protein molecule that were in a proper geometry for binding into two successive major grooves of DNA (Anderson et al., 1981). A model was proposed consistent with the protein structure and available chemical modification and biochemical data in which the two-fold symmetry axis of the protein coincided with that of the DNA and preserved two-fold symmetry in the protein-DNA complex (Ohlendorf et al., 1982). This initial model predicted essentially identical protein-DNA interactions and conformations for the protein and DNA in both halves of the protein-DNA complex.

The basic features of this model, in so far that two recognition helices hydrogen-bond and make van der Waals contacts within two successive major grooves of B-form DNA have been verified by the structural work on protein-DNA complexes of related proteins (Wolberger et al., 1988; Aggarwal et al., 1988; Jordan & Pabo, 1988; Otwinoski et al., 1988; Lamerichs et al., 1989) and have been reported in a preliminary account of the structure for a Cro protein-operator DNA complex (Brennan & Matthews, 1989). The complexes retain the general two-fold symmetry properties of the unbound DNA and protein. However, none of the six Cro and λ repressor binding sites in the O_L and O_R DNA regions of phage λ DNA are perfectly symmetric[1]. For binding to λ repressor, the asymmetry of the O_L1 and O_R1 operator DNA sequences is sufficient to effect different results for base-pair substitutions at symmetry related positions in the operator DNA (Sarai & Takeda, 1989) and to observe slightly different conformations of the protein in each half of the protein-DNA complex (Jordan & Pabo, 1988).

---

[1] The phage λ O_L and O_R operators cannot ever be truly symmetric since they contain an odd number of base-pairs. With the same sequence on either side of the central base-pair, an operator composed of an odd number of base-pairs could be functionally symmetric if the central base-pair played no role in the structure of the free or complexed DNA. However, this is unlikely since from known structures of DNA (e. g. Chapter IV; Dickerson, 1983) the conformation is observed to be dependent on base-sequence, which therefore destroys the symmetry of the full operator, especially for the base-pairs immediately adjacent to the central base-pair.

In this chapter, it is asked how the base-pair differences between the two halves of the O_R3 operator affect the binding of Cro repressor protein. Cro repressor protein is titrated with a right half-operator, termed R9, which is different from the L10 half-operator (Chapters IV and V; Figure VI-1). Binding strengths and the conformation of the Cro repressor are compared in the two protein-DNA complexes. Differences between the Cro-R9 and Cro-L10 protein-DNA complexes suggest some similarities to aspects of asymmetry found in the λ repressor-DNA complex. Mechanisms for the recognition of the Cro transcriptional regulatory protein have considered only interactions between a single Cro monomer and a consensus half-operator site, with the assumption that the interactions in the remaining half-site are related by the two fold symmetry of the complex. A revised model is suggested which allows asymmetry in the two halves of the protein-DNA complex.

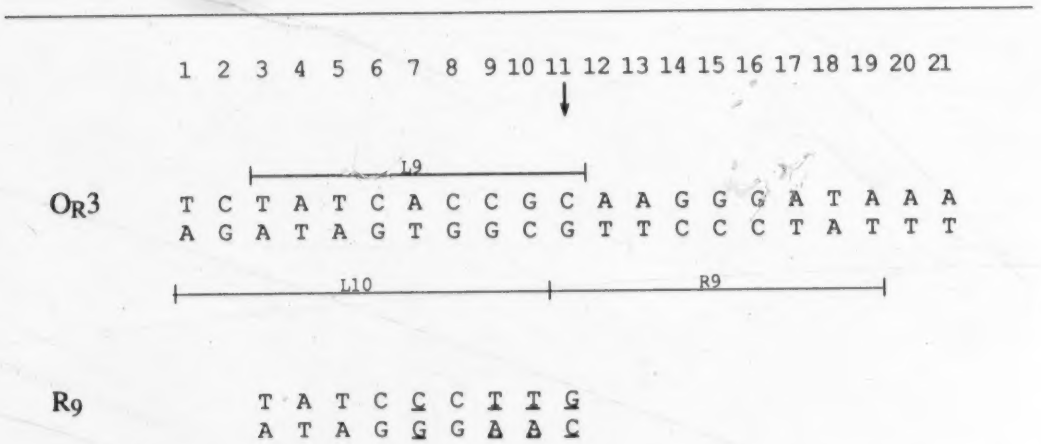

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
L9
OR3
T C T A T C A C C G C A A G G G A T A A A
A G A T A G T G G C G T T C C C T A T T T
L10
R9
R9
T A T C C C T T G
A T A G G G A A C

Figure VI-1 DNA half-operator sequences

Illustrated are DNA half-operator sequences from the O_R3 operator of phage λ. The arrow indicates the center of approximate two fold symmetry. In the lower part of the figure, the R9 sequence has been rotated to shows the correspondence to the left half-operator. Base-pair changes from the left half are underlined. Base-pair composition corresponding to the center of the operator (at positions 10-12) is not expected to appreciably affect the affinity for Cro repressor protein (Figure I-12).

B. Experimental Procedures

Cro protein was prepared as discussed in Chapter V using the pTR214/K802 Cro expression system. The initial concentrations of proteins for both titrations with DNA were 75 μM. All protein concentrations quoted refer to the dimeric form of Cro. The concentration of R9 duplex DNA (gift from Dr. W. Anderson) was determined using the snake venom phosphodiesterase assay (Chapter IV). L10 DNA was prepared as indicated in Chapter IV. Stock solutions of DNA were 3-4 mM. Buffers used were 0.06 M KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, pH 7.1 (direct meter reading), 50 μM EDTA, in D_2O at 30°C.

Two titrations of the protein were performed by adding aliquots of DNA. The first titration used L10 DNA, the second, R9 DNA. NMR spectra were recorded at each titration point on a Varian XL-400 spectrometer, with an operational frequency of 400 MHz for protons. 2400 transients were summed into 8K data points with a spectral width of 4000 Hz and a delay time between scans of 1.5 seconds. Chemical shifts were measured relative to DSS.

C. Results

Figure VI-2 shows the aromatic region of the ^1H NMR spectrum for Cro repressor protein as DNA is titrated in. The resonance lines for His 35 H2 and Tyr 26 3,5 protons are indicated for the unbound protein. The chemical shift positions for these protons are given for the Cro-L10 DNA complex from Chapter V. The His 35 and Tyr 26 resonance lines behaved differently on the addition of R9 DNA and their *upfield* changes in chemical shift could be monitored as a function of DNA:protein ratio. These and other methyl proton changes in chemical shift are given in Table VI-1 and are shown combined in Figure VI-3. The addition of R9 past a 1:1 ratio of DNA to protein produced no significant changes in the NMR spectrum of the protein. The data is consistent with only one R9 half-operator binding with appreciable strength to the Cro dimer. One step binding of protein, P, and R9 DNA, R, to form a protein-DNA complex, PR, can be expressed as:

$$P + R \underset{k_d}{\overset{k_a}{\rightleftharpoons}} PR \qquad (1)$$

and the equilibrium dissociation constant, K_d, for this reaction is:

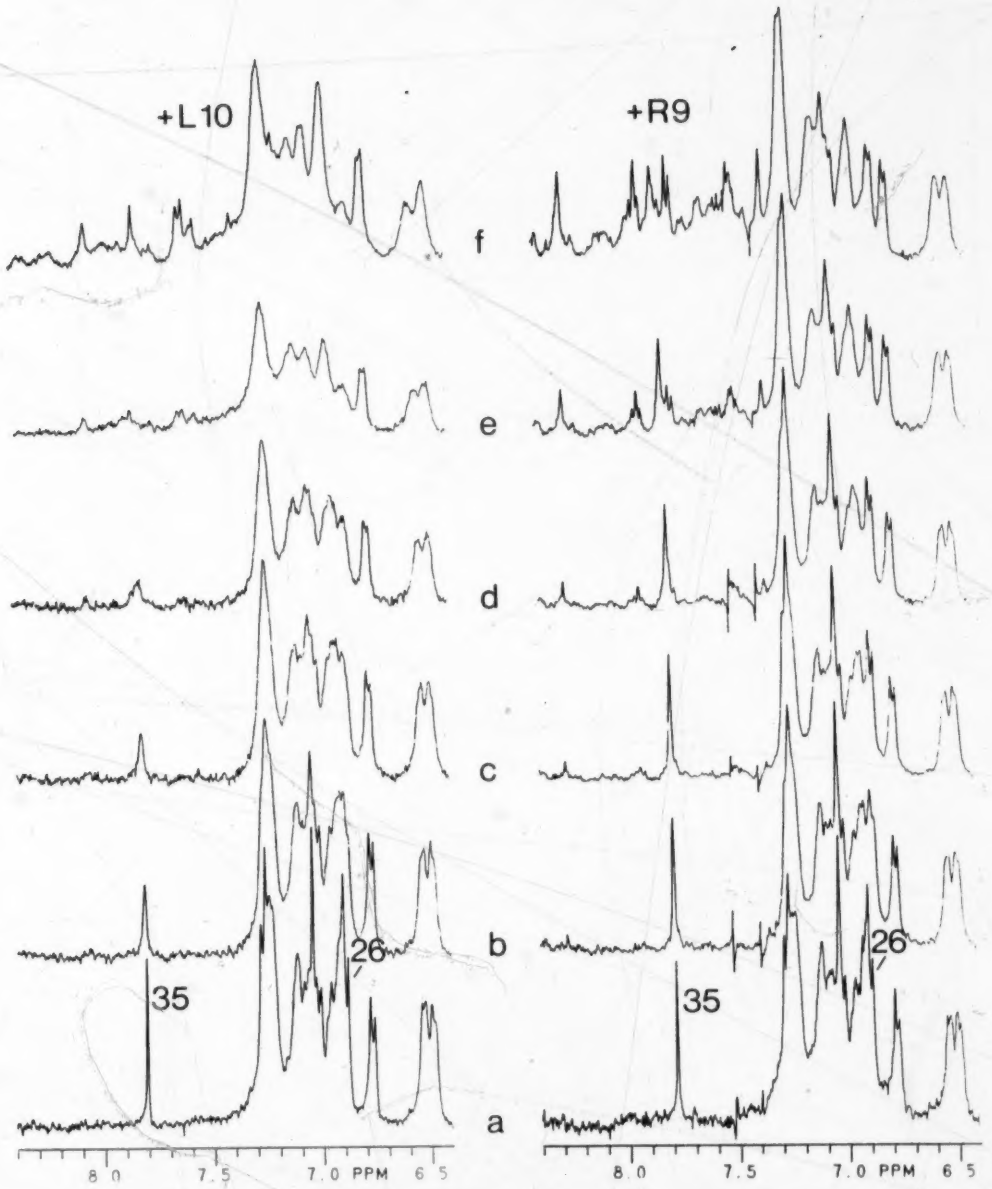


Figure VI-2 Titration of Cro aromatic side-chain ^{1}H NMR resonances with DNA
Titration of Cro repressor protein with a left half-operator, L10 DNA, and a right half-operator, R9 DNA. Shown are 6.4 to 8.4 ppm regions of ^{1}H NMR spectra taken at 30°C, in 0.06 M KCl, pH 7. On the left is the titration with L10 DNA, on the right is the titration with R9 DNA. The molar ratios of DNA to protein (dimer) are: (a), 0.00; (b), 0.19; (c), 0.38 for L10, 0.27 for R9; (d), 0.57; (e), 1.03; and (f), 2.0. The numbers indicate the resonance lines for the H2 proton of His 35 and the 3,5 protons of Tyr 26.

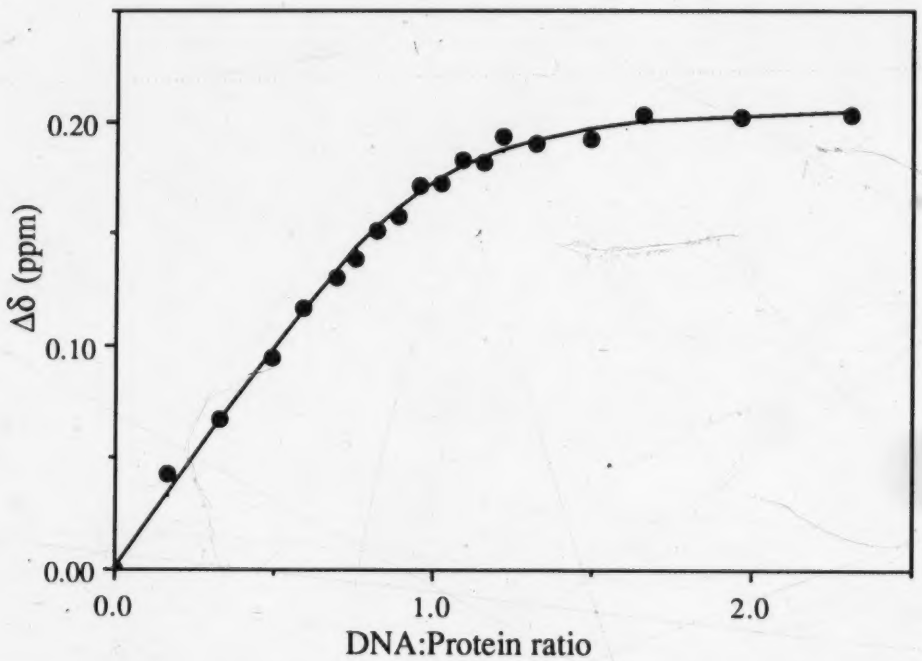


Figure VI-3 Titration shifts of Cro ^{1}H NMR resonances with R9 DNA

Titration shifts of ^{1}H NMR resonances of Cro repressor protein upon addition of the right half-operator DNA, R9. The line drawn through the points is calculated with an equilibrium dissociation constant of 3 μM. The initial protein concentration is 75 μM (dimer).

Table VI-1 Chemical shift changes in Cro repressor upon addition of DNA[a]

Nucleus	Position in Free Cro (ppm)	Change in Chemical Shift (ppm)			
		L10	R9	L9	O_R3
His 35 2H	7.78	+0.48	+0.025	+0.30	+0.32
Tyr 26 3,5	6.89	+0.35	- 0.088	>+0.1	>+0.1
Tyr 51 3,5	6.78	+0.02	0.00	0.00	0.00
Tyr 10 3,5	6.52	+0.10	0.00	0.00	+0.04
Phe 14 2,6	6.49	+0.01	0.00	-0.02	+0.01
Ala 29 CH_3[b]	1.61	[d]	+0.02	[d]	[d]
Ala 33 CH_3[b]	1.52	[d]	- 0.02	[d]	[d]
Thr 64,65 CH_3[c]	1.21	-0.05	0.00	[d]	[d]
Val 25 CH_3	0.63	-0.02	-0.02	[d]	[d]
Leu 42 $\delta 2$ CH_3	0.53	+0.01	0.00	0.00	+0.01
Ile 40 γ CH_3	0.31	- 0.02	0.00	- 0.03	- 0.04
Leu 42 $\delta 1$ CH_3	0.01	+0.05	0.00	+0.02	+0.04
Ile 40 δ CH_3	-0.26	+0.06	0.00	+0.11	+0.14

[a] Data for L9 is from Kirpichnikov et al. (1984) and data for O_R3 is compiled from Lee et al. (1987). The numbers indicate the observed change in chemical shift for the Cro:protein complexes at 20-22°, except for R9 (30°).

[b] Tentative assignment.

[c] The doublets of Thr 64 and Thr 65 methyl resonances overlap. The reported chemical shift is their center.

[d] No chemical shift change is available.

$$K_d = \frac{k_d}{k_a} = \frac{[P]\,[R]}{[PR]} \qquad (2)$$

With $P_o = [P] + [PR]$ and $R_o = [R] + [PR]$, K_d is:

$$K_d = \frac{\{P_o - [PR]\}\,\{R_o - [PR]\}}{[PR]} \qquad (3)$$

Since the chemical shift changes are in fast exchange, $\delta_{obs} = \delta_f f_f + \delta_b f_b$, where the symbols are as defined in Chapter V. The observed change in the observed chemical shift, $\Delta\delta$, is:

$$\Delta\delta = \frac{[PR]}{P_o} (\delta_b - \delta_f) \qquad (4)$$

As described in Chapter V, the titration data was fit by adjusting K_d and $(\delta_b - \delta_f)$, and solving equations (3) and (4) with a quadratic function. The equilibrium dissociation constant calculated for the binding of R9 to Cro is 3±1 μM.

A more complex situation arose when titrating Cro with L10 DNA. At DNA:protein (dimer) ratios of greater than 0.3, the solution became cloudy. Maximum precipitate was noted at a DNA:protein ratio of 1.0. At a DNA:protein ratio of approximately 1.5, the solution turbidity was noted to diminish. At a DNA:protein ratios close to 2.0, the solution was clear. Note the increase in NMR signal for the 2 to 1 complex over the 1 to 1 complex. This suggests that at these ionic strengths the 1:1 L10:Cro complex is insoluble and that the 2:1 $[L10]_2$:Cro form is soluble and that, of course, two L10 DNA molecules bind to the Cro dimer. What is observed in the NMR spectrum is mostly the free protein and the $[L10]_2$:Cro species, and only a small amount of the 1:1 DNA:protein complex.

Adding more complexity to the situation is that any observable titration shifts for L10 would probably be in slow or intermediate exchange. The R9 DNA sequence contains a few non-consensus base-pairs (Figure VI-1) that would be expected to decrease binding. Using the graph of Figure I-12, the binding of L10 DNA for Cro should be approximately 1.8 Kcal/mole better, which corresponds to a K_d of about 40 times stronger than R9, or about 0.1 μM. Assuming a association rate constant, k_a, of approximately $4 \times 10^8\ M^{-1}\ s^{-1}$ (corrected for temperature, from Kim et al., 1987), the dissociation rate constant would then be about $40\ s^{-1}$. Fast exchange would only be expected for resonances that shift by considerably less than $(k_d/2\pi)*400$ MHz ≈ 0.013 ppm. Since His 35 εH and Tyr 26 3,5 protons shift by about +0.3 ppm, they would be observed in slow exchange, if the 1:1 DNA protein complex were soluble.

D. Discussion

1. The asymmetry of O_R3 operator DNA

The stoichiometry of the binding of the half-operators to the Cro dimer is critically dependent on the presence of the central base-pair. Without this base-pair, two half-operators can bind to the Cro dimer, without any large cooperativity effects. If the middle base-pair is present, the binding of a second half-operator is severely hindered. 1:1 binding has also been observed by monitoring both the CD signal of the DNA and the ^1H NMR signal of the protein for binding of the analogous L9 half-operator DNA to Cro (Kirpichnikov et al., 1984; Kirpichnikov et al., 1985). Consistent with the half-operator binding data, insertion of an extra base-pair into the O_R3 sequence allows only non-specific binding of Cro to DNA but deletion of the central base-pair still allows specific binding, although affinity is decreased 1000-fold (Takeda et al., 1989).

The Tyr 26 3,5 protons of Cro shift downfield 0.23 ppm when forming the $[L10]_2$:Cro complex but shift upfield 0.088 ppm in the R9:Cro complex (Table VI-1). These observed chemical shift changes are presumably an average between the signal from the Cro monomer without bound DNA and the signal from the Cro monomer complexed to R9 DNA. The Ohlendorf et al. (1982) model predicts that the two Tyr 26 residues of the Cro dimer make sequence-specific hydrogen contacts to the thymine O4 oxygens of base-pairs 3 and 19 (Figure VI-1). The systematic base-substitution experiments of Sarai & Takeda (1989) suggest additional hydrogen bonds to the thymine oxygens at base-pairs 4 and 18 (one of which is presumably water mediated). The results presented here argue against the same interactions in both halves of the full operator.

In the 17 base-pair O_R3 operator sequence, the last four base-pairs are related exactly by a two-fold symmetry axis (corresponding to base-pairs 3-6 and 16-19 in Figure VI-1, respectively). The assignment of the ^1H NMR spectrum of O_R3 shows the same chemical shifts for the outer two symmetry related base-pairs (within ±0.03 ppm; Wemmer et al., 1984; Hahn et al., 1985). Likewise, in the 17 base-pair O_R3 operator for which thymine is replaced by 5-fluorouracil, the fluorines in symmetry related pairs have the same chemical shift in the absence of Cro protein (Metzler & Lu, 1989). However, when Cro is added the chemical shifts of symmetry related fluorines differ. The fluorine in base-pair 3 (replacing the methyl of the thymine) shifts downfield about 0.3 ppm but the symmetry-related fluorine in base-pair 19 does not shift. In addition, the fluorine in base-pair 4 shifts about 0.1 ppm upfield but the fluorine in base-pair 18 shifts 0.3 ppm upfield. The shifts for the adjacent fluorines towards the center of O_R3 (base-pairs 5 and 17) are the same. Therefore,

there is an asymmetry in the environment at the two ends of the operator, despite the fact that the base-pairs involved and the Cro repressor dimer are symmetric. This asymmetry is reflected in the different behavior of the Tyr 3,5 protons on complex formation with the different half-operators. The differential chemical shifts of Tyr 26 on binding the left and right half-operators is not due to the extra two base-pairs at base-pair positions 1 and 2 of L10 which are missing in R9. A similar titration has been made with a left-half L9 (see Figure VI-1) operator DNA for which the same chemical shift changes as in the titration with L10 are observed for the Tyr 26 3,5 protons and His 35 2H upon complex formation with Cro repressor (Kirpichnikov et al., 1984).

Likewise, the His 35 2H ^{1}H NMR signal behaves differently when the two different DNA duplexes are added. When L10 DNA is added the His 35 2H reappears at 0.48 ppm further downfield at 8.24 ppm. Kirpichnikov et al. (1984) have observed the same shift in the L9:Cro complex, and have shown that the change mainly results from a different pKa for the histidine side-chain. The ionization constant increases from about 6.3 to 6.9, indicating de-protonization of the histidine, consistent with the Ohlendorf et al. (1982) model which places the His 35 side-chain in the vicinity of a phosphate, which is likely the phosphate between base-pair 6 and 7 or between 7 and 8 of the upper strand in the consensus half of the O_R3 operator (Figure VI-1). A downfield shift of less than 0.1 ppm was noted for the His 35 2H chemical shifts corresponding to the protonated and deprotonated states. In the titration with R9 DNA, the His 35 H2 proton shifts downfield 0.025 ppm. Therefore, in the non-consensus right-half of the operator, this histidine does not appear to be interacting with the phosphate. In the right half-operator base-pair 7 is a CG base-pair, not an AT base-pair as in the left or consensus half-operator. Also base-pair 9 is an AT, whereas it is a GC in the left half. The different base-pair compositions at positions 10 and 11 likely play only a small role in the binding of Cro to operator DNA (Sarai & Takeda, 1989). Because key base-pairs are different in the right half-operator, the position of the recognition helix of Cro is likely changed, resulting in His 35 and Tyr 26 near the ends of the helix making different interactions with the DNA.

Closer inspection of the base-substitution experiments of Takeda et al. (1989) also reveals asymmetry for the binding of Cro to the O_R1 operator, and by analogy, the O_R3 operator. Figure I-12 shows the changes in free energy upon base-substitution for only the consensus half of the O_R1 operator (which has the same sequence as base-pairs 3-10 of L10). Changes for the base-pairs related by two-fold symmetry are similar, but not quite the same. For example, the base-pair at positions 4 and the symmetry-related position 18 of the O_R1 operator is T:A. The initial model by Ohlendorf et al. (1982) would predict the same interactions between each half of the Cro dimer and operator DNA, and therefore the

same loss of binding upon replacement by base-pair substitution in each side of the full operator. This is not the case. Replacement of the T:A at position 4 by any other base-pair results in about a 240-fold weaker equilibrium dissociation binding constant or about 3 Kcal•mol^{-1} loss in free binding energy. Replacement of the T:A at position 18 results in only 40-fold weaker binding, or about a 2 Kcal•mol^{-1} loss in free binding energy. Therefore, although Cro is two-fold symmetric in the absence of DNA, the protein is not making exactly the same interactions in both sides of the full operator.

The upfield shifts of Tyr 26 3,5 protons are also observed for the interaction of the 26-39 fragment of Cro with a non-operator DNA decamer, suggesting that the R9 DNA:Cro complex is non-specific (Mayer et al., 1983). However, if non-specific DNA binding were occurring, it should be possible for two R9 DNA molecules to bind to the Cro dimer. Stoichiometric 1:1 binding is observed for the interaction of R9 DNA and Cro, as is observed with the analogous L9 DNA. Secondly, the non-specific binding equilibrium dissociation constant for the binding of full operator DNA to Cro is estimated to be about 5 x 10^{-8} M (Takeda et al., 1986; Takeda et al., 1989). This corresponds to a free energy change of about -10 Kcal•mol^{-1}. Assuming about one half the interaction energy for the interaction of a half-operator to Cro, the K_d should be about 2 x 10^{-4} M. The observed binding is a hundred fold stronger.

The temperatures used in this study may be sufficiently high to cause thermal denaturation for a significant population of the DNA. However, the melting temperatures of the R9 and L10 sequences (about 40°C in 0.06 M KCl) should not be very different so that their binding to Cro should remain comparable. More importantly, the denaturation temperature of Cro protein is about 45° at this ionic strength (42°C in 20 mM phosphate buffer, Iwahashi et al., 1982; 47°C in 0.2 M KCl, Chapter V) and denaturation temperatures of protein-nucleic acid complexes are usually higher than that of their components. The complexes of Cro and half-operator DNA are therefore expected to be stable[2] at 30°C.

The ^{1}H NMR resonances of methyl groups of Cro also indicate a difference in the binding of the left and right operators (Figure VI-4). The Thr 64, Thr 65, Leu 42 and Ile 40 side-chain methyl group chemical shifts all change for the addition of L10 DNA, but not for the addition of R9 DNA. Some chemical shift changes are the same however, such as

---

[2] A preliminary titration of Cro protein was performed with R9 DNA at 15° in a buffer containing 0.08 KCl, 10 mM KH_2PO_4, 10 mM K_2HPO_4, pH 7.0±0.1, 500 μM EDTA. The equilibrium dissociation binding constant was 14±10 μM, but more importantly, the same chemical shift changes are observed in the protein at this lower temperature.

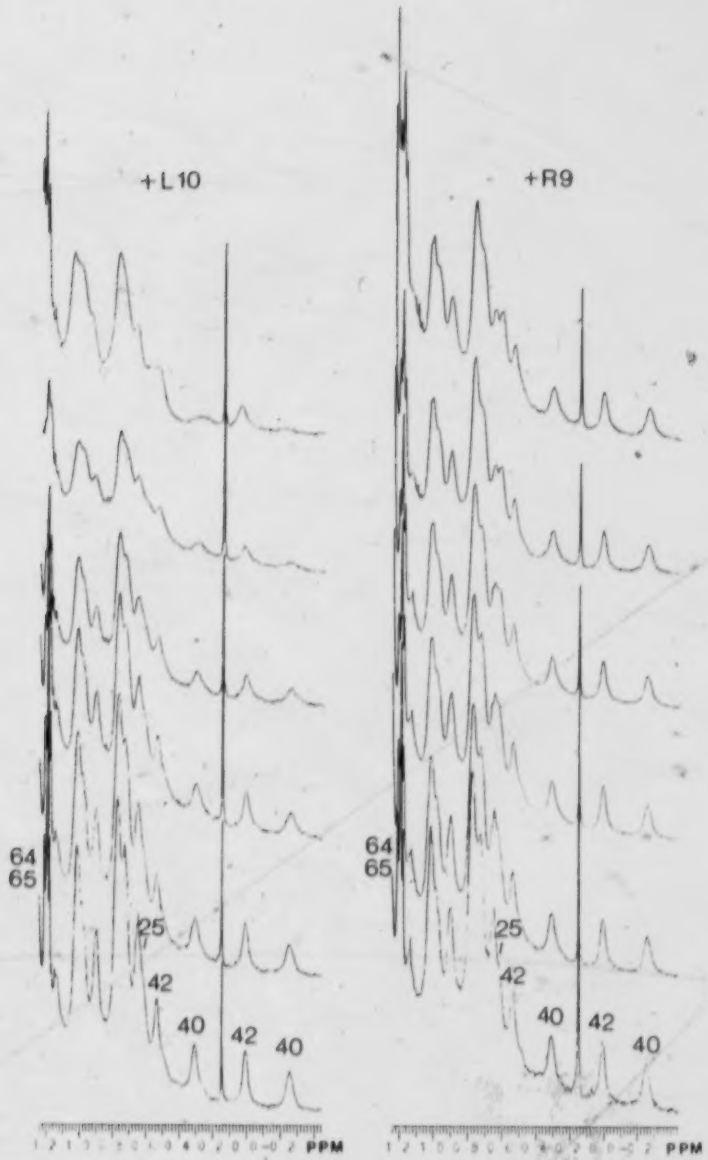


Figure VI-4 Titration of Cro methyl ^{1}H NMR resonances with DNA

Titration of Cro repressor protein with a left half-operator, L10, and a right half-operator, R9. Shown are the -0.45 to 1.25 ppm regions of ^{1}H NMR spectra taken at 30°C, in 0.06 M KCl, pH 7. On the left is the titration with L10 DNA, on the right is the titration with R9 DNA. The molar ratios of DNA to protein (dimer) are: (a), 0.00; (b), 0.19; (c), 0.38 for L10, 0.27 for R9; (d), 0.57; (e), 1.03; and (f), 2.0. The numbers indicate the assignment of resonance lines to specific residues of the protein.

observed shift for the methyl group at 0.63 ppm (Val 25), indicating that the R9 and L10 DNA do have some of the same interactions between Cro and DNA.

Although the Cro repressor protein in the absence of DNA is symmetric, its major binding site on phage λ DNA is not. Correspondingly, the protein:DNA complex is not perfectly symmetric which results in different interactions between the Cro monomer and left-half operator DNA from those between Cro monomer and right-half operator DNA.

2. Model for Cro-DNA aggregation

In solutions of ionic strengths of approximately 0.2 M, the 1:1 L10:Cro dimer complex did not result in the formation of a precipitate, despite the titration being performed at 20 fold higher protein and DNA concentrations (Chapter V). A model can be proposed to explain the aggregation phenomena of Cro-nucleic acid complexes (Figure VI-5) based on the observations presented in this chapter for the interaction of half-operators with Cro, and observations that non-specific DNA interactions are favoured by low ionic strengths (Boschelli, 1982). In the 1:1 complex, one site of the protein is still available for binding to DNA. In this model, the unoccupied Cro monomer binds non-specifically to DNA already present in a protein:DNA complex. Furthermore, it binds mainly to the back-side of the consensus part of the DNA (base-pairs 3-8) and to the two base-pair extension of the half-operator (base-pairs 1 and 2 of Figure VI-1). When both sites of Cro dimer are occupied, as in the $[L10]_2$:Cro complex or the O_R3:Cro complex, aggregation cannot occur as readily. Likewise, when the ionic strength of the solution is high, extensive non-specific interaction between DNA and protein, and the resulting precipitation of DNA-protein complexes does not occur.

The precipitation event should also take into account the net ionic strengths of the Cro and DNA. The fully charged amino acid residues of a Cro monomer are 8 lysines, 3 arginines, 3 glutamates and 3 aspartates. In addition, the N-terminal amino group and His 35 are partially protonated. The net charge of a Cro dimer is therefore about +12. L10 DNA has 18 phosphate groups (terminal ends are hydroxylated) and therefore a charge of -18. R9 DNA has a net charge of -16. Although the net charge of the protein:DNA complex can be only estimated since all species are surrounded by a excess of counter-ions (K^+, Cl^-, PO_4^-) which are displaced from the surface of the protein and DNA on their interaction and it is difficult to predict the effect on charge by the interaction of non-charged groups with charged ones, the net charge on both half-operator:Cro and $[half\text{-}operator]_2$:Cro complexes is probably sufficiently negative to avoid precipitation by being at an isoelectric point. In any case an ionic charge explanation is insufficient to account for the differences in solubility between the R9:Cro and L10:Cro complexes, since both should carry net negative charge, with the L10:Cro complex being of higher net charge.

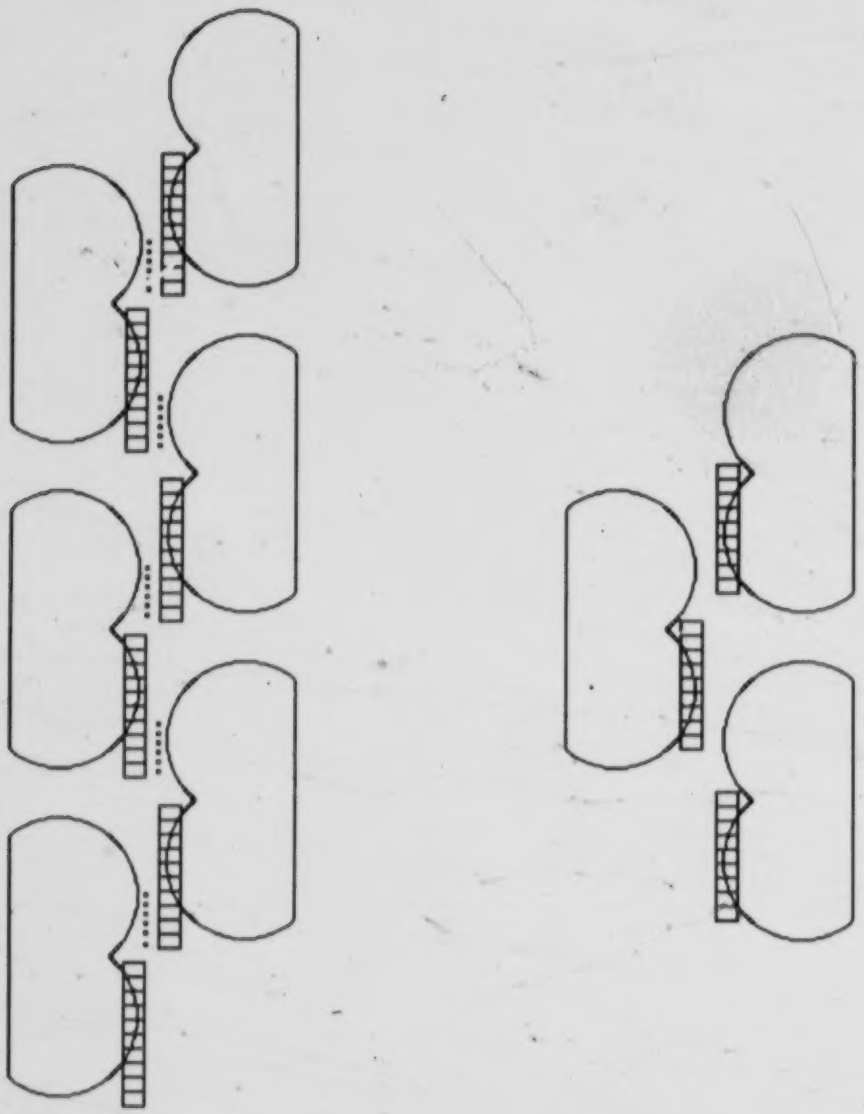


Figure VI-5 Model for the aggregation of 1:1 L10 DNA:Cro protein complexes.
Shown on the left is a model for the aggregation of the 1:1 protein:DNA complex of L10 DNA and Cro repressor. The Cro dimer is represented by the two-fold symmetric protein globule and L10 DNA is represented by the rectangle divided into 10 parts representing the ten base-pairs of the DNA. Specific interaction between protein and DNA is indicated by their close association. Extensive non-specific interactions (• • • •) lead to precipitation. On the right are shown 1:1 DNA:protein complexes with R9 DNA. Non-specific interactions do not form as readily as with the 1:1 L10:Cro complex.

R9 DNA:Cro complexes do not form protein:DNA precipitates since the R9 DNA does not have the corresponding two base-pair extension and the binding of one R9 DNA to the Cro dimer partially occludes the second binding site. L9 DNA should, by this model, display the same solubility characteristics as R9. Although L9 DNA Cro formation was coincident with additional broadening of all spectral lines because of non-specific interaction, no precipitation of the L9:Cro complex was noted (Kirpichnikov, 1984).

3. The interaction of Cro with O_R3 operator DNA

Some caution should be exercised in extending the results of titrations with left and right half-operators too literally to the interaction of Cro with the full operator. Exactly the same complex as with the full operator would not be formed if a left and a right operator were added simultaneously to the Cro dimer as they contain a double strand break at the center of the operator. Although the identity of the central base-pair is not very important for the binding of Cro to operator DNA (Takeda et al., 1989), the preliminary crystallographic determination of the Cro-O_R3 operator complex indicates a change in DNA conformation at the center of the operator (Brennan & Matthews, 1989), as is also indicated by the large chemical shift change for the imino proton of the central base-pair (Kirpichnikov et al., 1984). Moreover, the available ^1H NMR chemical shift changes of the protein in the O_R3 complex indicate a conformation more like that of the left half-operator protein:DNA complex, and not the intermediate (if observed in fast exchange) expected for the average of the R9:Cro and L10:Cro complexes.

Since the chemical shift changes of the protein on interaction with left-half operator DNA are larger, and the binding strength stronger, than on interaction with right-half operator, a revised model for the binding of Cro to operator DNA would include asymmetry in the two halves of the operator. The left half is more tightly associated with Cro than the right half, with greater distortions in structure from that of the unbound protein. A similar non-symmetric situation is seen in the crystal structure of the λ repressor N-terminal DNA-binding domain complexed with full operator DNA (Jordan & Pabo, 1988; Sarai & Takeda, 1989). As discussed in Chapter 1, the asymmetry with λ repressor is more severe than in Cro. Base-substitution experiments with Cro give more or less symmetric results (Takeda et al., 1989) but are very asymmetric with λ repressor. The asymmetry with λ repressor probably results from monomer-monomer contacts preventing the same conformation of the protein when binding simultaneously to the two half-sites presented in a full operator. In Cro, the asymmetry is more subtle and is caused mainly by differences in base-pair composition between the two half-sites.

E. References

Aggarwal, A. K., Rodgers, D., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899-907.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754-758.

Boschelli, F. (1982) *J. Mol. Biol. 162*, 267-282.

Brennan, R. G., & Matthews, B. W. (1989) *Trends Biochem. Sci. 14*, 286-290.

Dickerson, R. E. (1983) *Sci. Am.* pp. 94-111.

Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Skryabin, K. G., & Kirpichnikov, M. D. (1985) *Eur. Biophys. J. 12*, 87-95.

Iwahashi, H., Akutsu, H., Kobayashi, Y., Kyogoku, Y., Ono, T., Koga, H., & Horiuchi, T. (1982) *J. Biochem. (Tokyo) 91*, 1213-1221.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893-899.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149-158.

Kirpichnikov, M. P., Kurochkin, A. V., Chernov, B. K., & Skryabin, K. G. (1984) *FEBS Letters 175*, 317-320.

Kirpichnikov, M. P., Yartzev, A. P., Minchenkova, L. E., Chernov, B. K., & Ivanov, V. I. (1985) *J. Biol. Struct. Dynam. 3*, 529-536.

Mayer, R., Lancelot, G., & Hélène, C. (1983) *FEBS Letters 153*, 339-344.

Metzler, W. J., & Lu, P. (1989) *J. Mol. Biol. 205*, 149-164.

Lamerichs R. M. J. M., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Rüterjans, H. (1989) *Biochemistry 28*, 2985-2991.

Lee, S. J., Shirakawa, M., Akutsu, H., Kyogoku, Y., Shiraishi, M., Kitano, K., Shin, M., Ohtsuka, E., & Ikehara, M. (1987) *EMBO J. 6*, 1129-1135.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718-723.

Otwinowski, Z., Schevitz, R. W., Zhang R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321-329.

Pabo, C. O., & Sauer, R. T. (1984) *Ann. Rev. Biochem. 53*, 293-321.

Sarai, A., & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. 86*, 6513-6517.

Schlief, R. (1988) *Science 242*, 893-907.

Takeda, Y., Kim, J. G., Caday, C. G., Steers, E., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1986) *J. Biol. Chem. 261*, 8608-8616.

Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. 86*, 439-443.

Wemmer, D. E., Shou, S.-H., & Reid, B. R. (1984) *J. Mol. Biol. 180*, 41-60.
Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature 335*, 789-795.

Chapter VII

The Structural Basis of Operator DNA-Cro Repressor Protein Recognition

A. Summary

This thesis has described the structural basis of the Cro repressor protein-O_R3 operator interaction from bacteriophage λ. The major technique used here to investigate the mechanisms by which protein and DNA recognize each other has been nuclear magnetic resonance spectroscopy.

In Chapter II, methods are given for improving the accuracy of macromolecular structure determination in solution using NMR techniques. The extrapolation of derived interproton distances to initial NOE mixing times is convenient and reliable. The major error in all distance determination methods arises through relayed NOE effects. The NOE-based refinement technique circumvents this problem. A proposal was made for the use of an NOE R factor to gauge the fit of observed NOE data to the experimentally determined structure. The NOE-based refinement technique allowed for local correlation time reduction factors analogous to the incorporation of thermal factors used in X-ray crystallography and mainly reflect the internal motions of macromolecules.

In Chapters III and IV, these methods are applied to three DNA duplexes. DNA octamers of alternating purine-pyrimidine sequence, $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$ and $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$, adopted a B-DNA type conformation in solution in the absence of external forces such as proteins being bound. The function of purine-pyrimidine sequences in biology was suggested by the observed base-stacking pattern to originate from dinucleotide structural units. The structure of the left ten base-pairs of the O_R3 operator from bacteriophage λ, d(T-C-T-A-T-C-A-C-C-G)•d(C-G-G-T-G-A-T-A-G-A), L10, was also determined to be B-type DNA.

At first it may seem that the NMR techniques were incapable of generating anything but a rather ordinary B-type DNA structures. Much of the biology of the three DNA duplexes could be well explained with Arnott & Hukins (1972) B DNA model, or more accurately, with the average features observed in one the first B DNA crystal structures (Dickerson & Drew, 1981). Recently, NMR studies of certain DNA sequences were shown not to be in regular conformation (e. g., the Bcl1 restriction endonuclease recognition site, Banks et al., 1989; bent DNA, Nadeau & Caruthers, 1989). But unlike the sequences studied here, these DNA molecules exhibited both unusual ^1H NMR chemical shifts and NOESY cross-peaks. The DNA studied in this thesis have more usual NMR parameters and just have sequences which adopt regular B type conformation.

The confirmation in the validity of the NMR methodology used here comes with the comparison of the L10 sequence to the corresponding base-pairs of a DNA-protein complex determined by established X-ray crystallographic techniques (Jordan & Pabo,

1988). The protein here was not Cro, but the N-terminal domain of a related protein, the λ or cI repressor. Both the λ and Cro repressor proteins recognize and bind preferentially to the same six sites on bacteriophage λ DNA. Their differing affinities for these six sites regulate the life cycle of the phage. There was good agreement between the DNA structure determined using NMR techniques and that using crystallographic techniques, except in regions of the DNA contacted by the protein. Therefore, the result indicates that not only can NMR be used to determine structures in solution with high precision, but for the transcriptional regulatory proteins of phage λ, the DNA plays a rather passive role. For the interaction of a protein monomer with the site present on one half of the operator, the DNA mainly presents a pattern of hydrogen bond donors and acceptors and van der Waals contacts to be read by the recognition α helix of the protein.

In Chapters V and VI studies were undertaken of the interaction of the Cro repressor with L10 and other DNA sequences. The presence of a disulfide cross-link in a site-specific mutant of Cro does not appreciably alter its structure, but greatly increases the thermal stability of the protein and its resistance to amide proton exchange by deuterium. The less flexible protein has a five-fold loss in affinity for L10 half-operator DNA. The results suggest that flexibility is not only required for dimeric Cro to simultaneously bind to both DNA-binding sites presented in each half of the full O_R3 operator, but also for the monomer subunit to undergo the conformational changes required to bind DNA.

The ^1H NMR resonances of DNA broaden differentially on adding Cro protein. The degree of line-width broadening is shown to correspond to the contacts made to the DNA by protein consistent with the Ohlendorf et al. (1982) model. Experimental conditions needed to study protein-nucleic acid complexes using ^1H NMR techniques include ensuring that the interaction of protein and DNA are well defined within NMR chemical exchange regimes and that the protein:DNA complex remains soluble. Particularly useful are higher ionic strengths to force the system to display fast exchange behavior and keeping concentrations of protein and DNA low. A model is presented for the aggregation of Cro:DNA complexes which will help to avoid solubility problems in the future. The binding strengths and stoichiometries of the interactions for the left and a right half-operator DNA with Cro repressor were measured. Small deviations from true symmetry for both halves of the O_R3 operator result in different protein conformations in each half of the Cro dimer:O_R3 complex.

The results of this thesis suggest several aspects of the recognition of DNA and protein that would be useful for design of mutant transcription regulatory proteins to have a desired characteristic or to understand a particular detail in the interaction between protein and DNA: (1), the DNA has a B type conformation, at least for the consensus half-operators in

the regulatory region of phage λ DNA; (2), flexibility in the protein is necessary for DNA binding, although a disulfide cross-link could be introduced into a repressor to allow the protein to work at a higher temperature (under oxidizing conditions) than the original; and (3), although a repressor protein may be two-fold symmetric in the absence of DNA, and the DNA may be nearly symmetric in its base sequence, a few base-pair changes between the halves of the DNA-binding protein site may be sufficient to cause different interactions in each half of the DNA:protein complex.

Recognition between the correct DNA operator site and the Cro repressor protein is initiated by Cro first binding to the B type DNA, and since the number of possible non-operator sites is much larger than the number of operator sites, at a non-operator DNA sequence. When bound to non-specific DNA, the Cro exists in an open conformation capable of reading the pattern of hydrogen-bonding and van der Waals contacts presented in major groove of the DNA (Takeda et al., 1986) and preferentially moves in a direction along the DNA (Kim et al., 1987), although occasionally dissociating from it. Whenever a large number of favorable contacts can occur between Cro and DNA, the association between protein and DNA tightens and the dissociation rate decreases. With the operator DNA sequence, an optimal number of hydrogen bonds and van der Waals contacts can form between the amino acid side-chains of the recognition helix of the monomer in each half of the Cro dimer with the base-pairs in each half of the operator DNA, thus forming a functional DNA-protein complex. The steric fit and conformation and flexibility of protein and DNA are binding modulators of the hydrogen-bond matrix that lays the foundation for specific recognition between Cro and operator DNA (Berg & von Hippel, 1988). The operator DNA may bend at the center, but generally retains a B DNA structure in each half of the nearly two-fold symmetric sequence. The flexible Cro protein also undergoes conformational change for optimal interaction with operator DNA. Since the base-pair sequence is not precisely the same in each half of the operator, the protein adopts different conformations in each half of the protein dimer-operator DNA complex.

The described improvements in ^{1}H NMR methodology to determine macromolecular structure in solution are of greater general significance for contributing to the field of biochemistry. Future structure determinations will circumvent the approximate distance determination procedures entirely by directly using the comparison between experimentally determined NOE intensities and the calculated values. Further progress will include a more precise description of the dynamics of these molecules. To date, the chemical shift has not been used for detailed structure determinations. A sufficient empirical data basis and computing resources are now beginning to be available to allow the incorporation of chemical shift values for structure determination. Better accuracy in the force field used for

restrained molecular dynamics calculations would also improve NMR structure determinations. For nucleic acids, additional experimental data is required to give a better description of the conformation about the phosphate. This would likely incorporate the stereospecific assignment of the 5' and 5" protons, with strong coupling effects taken into account. Assignment of the ^{31}P NMR resonance signals would yield more information (see below), probably by measuring coupling constants to protons of the DNA backbone. Incorporation of other heteronuclei (^{13}C, ^{15}N, ^{17}O, ^{19}F) at defined positions in the DNA or protein are being used for spectral simplification, although the proton will remain, at least in the short term, the winning nucleus for structure determination.

B. Future studies of protein:DNA recognition using NMR spectroscopy

In this section, I shall present some NMR observations relating to protein:DNA recognition not yet discussed in this thesis. These studies are either of a preliminary nature or are of limited scope. Experiments necessary for completion are briefly described.

1. Amide exchange studies of Cro repressor

Despite indications to the contrary (Weber et al., 1985), amide exchange measurements with Cro repressor are possible. Amide proton exchange has been measured at 10°C with one-dimensional ^{1}H NMR spectra. Only two resonances, Leu 42 and Glu 54 are fully resolved in one-dimensional experiments. Two-dimensional experiments will be necessary to obtain accurate exchange rates for the remaining amide proton resonances. At 10° and pH of 4.8, Leu 42 and Glu 54 have exchange rates of 0.015 and 0.11 hr^{-1}, respectively. In a buffer containing 0.5 M KCl, Cro amide protons exchange about 4 times slower than at 0.06 M KCl, indicating that higher ionic strengths increase the stability of the protein. Future experiments would measure amide exchange of the protein while bound to DNA.

Amide exchange rates were also measured for wild type and V55C Cro proteins at 20°C. The direct comparison is difficult since the wild type protein exchanges very rapidly ($t_{1/2} \leq 5$ hrs), while the V55C Cro exchanges quite slowly ($t_{1/2} \geq 500$ hours). The exchange rates measureable from one-dimensional spectra are shown in Table VII-1.

Table VII-1 Amide proton exchange rates in Cro repressors

	Wild type	V55C	
Leu 42	100±11	2.6±0.1	(x 10^{-3} Hr^{-1})
Glu 54	490±40	0.2±0.1	
Lys 56	110±31	0.7±0.2	
Ile 34/ Ala 36	107±20	2.4±0.9	

Proteins were approximately 0.2 mM (dimer) in 0.2 M KCl, 10 mM KH_2PO_4, K_2HPO_4, 50 μM EDTA, pH 5, 20°C.

2. ^{31}P NMR of DNA

Figure VI-1 shows the 121 MHz ^{31}P NMR spectrum of L10 DNA. Assignment of these well resolved resonances would enable a detailed study of the interactions that Cro (and λ) repressors make with the phosphates of DNA, and a more precise description of the conformation of the DNA backbone. However, initial heteronuclear COSY experiments have been unsuccessful, probably since the line-widths of the resonances are large compared to the spin-spin coupling constant used to generate the COSY cross-peak. Preliminary ^{31}P NMR spectra of R9 DNA have also been obtained in the absence and in the presence of Cro protein. Although the signal to noise ratio was poor for the titration of the DNA with Cro, it could be observed that some resonance lines changed in chemical shift (fast exchange), and others lost intensity (intermediate or slow exchange).

3. Interaction of Cro repressor with other half-operators

The crude synthesis products of other half-operator sequences of the O_R3 operator have been obtained (from T. Atkinson and M. Smith, University of British Columbia). A DNA fragment corresponding to the right half version of L10 (R10) and L9 (see Chapter VI) have been purified using NACS-20 column chromatography procedures (described in Chapters III and IV). Preliminary ^1H NMR spectra have been taken. Insufficient quantities were made for an easy detailed structural analysis, but the amounts should be sufficient to

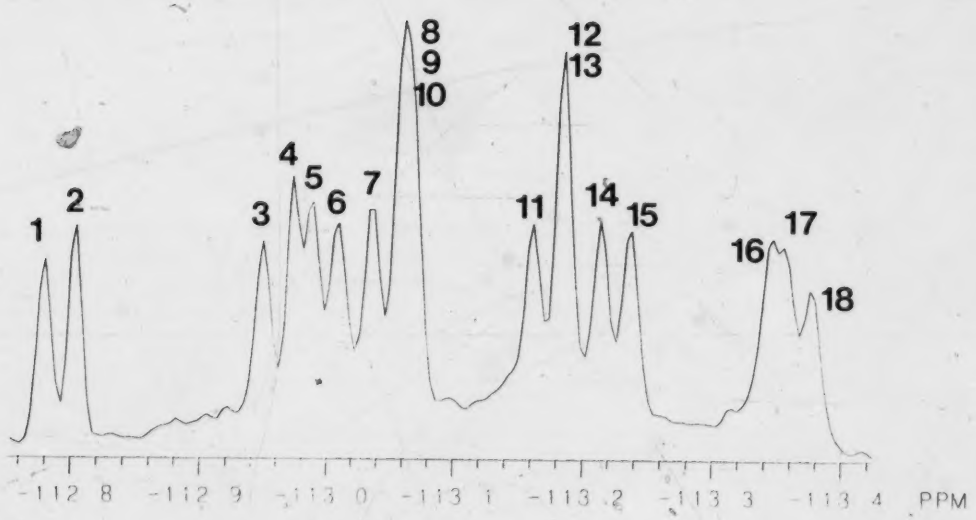


Figure VII-1 ^{31}P NMR spectrum of L10 DNA

The 121 MHz ^{31}P NMR spectrum of L10 DNA was taken at 30°C with proton decoupling. The sample was approximately 0.4 mM DNA in 50 mM KCl, 5 mM K_2HPO_4, pH 7.3, 0.3 mM EDTA buffer. 4096 transients were accumulated with an acquisition time of 2.0 seconds and a relaxation delay time of 0.5 seconds. The numbers count the number of phosphate resonances (L10 has 18 phosphates) and do not indicate assignment to a particular nucleotide. The resonance line for reference inorganic phosphate is at -110.6 ppm.

measure DNA binding to Cro. R10 should bind to the Cro dimer at a ratio of 2 to 1, display the same aggregation characteristics as L10, but have the binding strength of R9.

Other operator fragments are available in crude form. These include a nine base-pair segment encompassing the center of the operator and an R8 sequence, which is missing the outermost two base-pairs of R10. R8 should also bind at a 2 to 1 ratio, with the binding strength of R9, but without the two base-pair extension, the R8:Cro complex may not aggregate as strongly as the L10:Cro complex.

4. RNA polymerase tail peptide

The eukaryotic RNA polymerase II amino acid sequence contains a long repeating heptapeptide unit (-Ser-Pro-Ser-Tyr-Ser-Pro-Thr-)$_n$ at the carboxy terminus of the enzyme (Allison et al., 1985). Removal of this 'tail' sequence results in incorrect initiation of transcription. A peptide containing five of the repeat units has been characterized by ^{1}H NMR spectroscopy. Exchange rates and the chemical shift temperature dependence of the amide protons and are the same as unstructured peptides. Therefore, the amide protons of this segment do not participate in hydrogen bonding. The addition of monovalent cations (K^+), divalent cations (Mg^{2+}), or calf thymus DNA did not stabilize the structure. NOESY experiments were consistent with a unstructured conformation, with the ω angles of both proline *trans*. Either the repeating peptide needs to be in a larger protein context to adopt a defined conformation or the function of the tail requires it to be unstructured.

C. References

Allison, L. A., Moyle, M., Shales, M., & Ingles, C. J. (1985) *Cell 42*, 599-610.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504-1509.

Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6696-7010.

Berg, O., & von Hippel, P. H. (1988) *Trends Biochem. Sci. 13*, 207-211.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761-786.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893-899.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149-158.

Nadeau, J. G., & Caruthers, D. M. (1989) *Proc. Natl. Acad. Sci. 86*, 2622-2626.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) *Nature 298*, 718-723.

Takeda, Y., Kim, J. G., Caday, C. G., Steers, E., Jr., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1986) *J. Biol. Chem. 261*, 8608-8616.

Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985) *Biochemistry 24*, 4553-4562.

Appendix 1 FORTRAN code for NOE-based structure refinement

This code was added to the energy minimization section of the GROMOS molecular simulation package. This code is also applicable for molecular dynamics calculations.

A. Common block, 'em.common.2'

```
      COMMON /NOER/ ILDR(220),NILDR,PM(250),NPRFL(250),
     2              TMIX,TAUC,RADIUS,VOMEGA,NPROT,
     3              IDR3(650),JDR3(650),OBS0(650),CNOE,
     4              DBYD(1918),NOEDV,TMIX2,NNR1,NNR2,
     5              NOM,MGRP(5),MGRPC(5)
```

B. Changes made to the main energy minimization program to read in NOE files, find the non-exchangeable protons from the coordinates and molecular topology program.

```
C
      INCLUDE 'em.common.2'
C
      WRITE(6,28) 'Noe restraint file (mix time 1).....: ',FORT20
      WRITE(6,28) 'Noe restraint file (mix time 2).....: ',FORT21
   28 FORMAT(A41,A31)
C
C*****
C     READ IN NOE RESTRAINT FILES (NO. OF NOE RES FILES=NNRF)
C     TAUC IN NANOSECONDS, RADIUS IN NANOMETERS,
C     NOEDV IS NOE DERIVATIVE PROCEDURE-CAN BE ZERO TO APPROX.-
C     OTHERWISE IS TYPICALLY SET TO 5 OR 10
C
      READ (4,39) NNRF,CNOE,RADIUS,TAUC,NOEDV,NOM
      IF(NOM.GT.0) READ(4,37) (MGRP(I),I=1,NOM)
      IF(NOM.GT.0) READ(4,37) (MGRPC(I),I=1,NOM)
C
   37 FORMAT (10I5)
C
   65 FORMAT (1X,I10,F10.3,2F10.5,2I10)
C
C*****READ COORDINATES, PRINT TITLE
      READ(18,40) TITLE1
      READ(18,92) NATOM
      IF (NATOM.LT.NR) GOTO 1034
      I3=0
      N=0
C     Set full NOE derivatives to zero
      DO 76 J=1,NR
      DO 777 M=1,3
      DBYD(I3+M)=0.0
  777 CONTINUE
C  Find non-exchangeable protons, read in correlation time
C  adjustment factor.
      IF(IAC(J).EQ.38) GO TO 78
      READ(18,92) KK,K,(X(I3+M),M=1,3)
      I3=I3+3
      GO TO 76
   78 N=N+1
```

```
      READ(18,92) KK,K,(X(I3+M),M=1,3),PM(N)
      IF(PM(N).EQ.0) PM(N)=1.000
      NPRFL(N)=J
      I3=I3+3
   76 CONTINUE
      NPROT=N
      GOTO 100
   92 FORMAT (I5,10X,I5,3F8.3,3F8.4)
C  Read in observed NOE intensities
C
  231 IF(NNRF.EQ.0) GO TO 261
      TAUC=TAUC*1.E-9
      READ(20,40) TITLE1
      READ(20,219) NNR1,TMIX,VOMEGA
      VOMEGA=VOMEGA*2.0E6*3.1415927
      I=1
      READ(20,228) IDR3(1),JDR3(1),OBS0(1)
      ILDR(1)=IDR3(1)
      DO 235 N=2,NNR1
      READ(20,228) IDR3(N),JDR3(N),OBS0(N)
      M=0
  240 M=M+1
      IF(IDR3(N).EQ.ILDR(M)) GO TO 235
      IF(M.NE.I) GO TO 240
      I=I+1
      ILDR(I)=IDR3(N)
  235 CONTINUE
C
C  Read in Second observed NOE intensities file, if necessary
      IF(NNRF.LT.2) GO TO 246
      WRITE(6,226)
      READ(21,40) TITLE1
      WRITE(6,31) TITLE1
      READ(21,219) NNR2,TMIX2
      WRITE(6,227) NNR2,TMIX2
      NNRTOT=NNR1+NNR2
      DO 245 N=NNR1+1,NNRTOT
      READ(21,228) IDR3(N),JDR3(N),OBS0(N)
      M=0
  250 M=M+1
      IF(IDR3(N).EQ.ILDR(M)) GO TO 245
      IF(M.NE.I) GO TO 250
      I=I+1
      ILDR(I)=IDR3(N)
  245 CONTINUE
  246 NILDR=I
      WRITE(6,9046) NILDR
 9046 FORMAT (' # OF PROTON SUBSETS WILL BE',I5)
  226 FORMAT(/,1X,' 9.   R E S T R A I N E D   N O E S          :')
  227 FORMAT(/,1X,I5,1X,'NOE RESTRAINTS AT MIX TIME ',F10.5,' SECS')
  228 FORMAT(2I5,F6.3)
```

C. NOE calculation subroutine

```
      SUBROUTINE NOERE (NB,IB,JB,NM,NAM,CB,B0,NTB,BOX,BETA,X,F,
     2                  EB,XB0,EB0,NSTEP,NOERPT)
C
CCCCCC J. BALEJA, EDMONTON (ADAPTED FROM DISRE  ROUTINE)  CCCCCCCCCCC
C                                                                      C
C     SUBROUTINE NOERE (NB,IB,JB,NM,NAM,CB,B0,NTB,BOX,BETA,X,F,        C
C                                                                      C
C    2                  EB,XB0,EB0,NSTEP,NOERPT)                       C
C                                                                      C
C    NOERE WILL SUPPLY THE DISTANCE RESTRAINT ENERGIES AND FORCES      C
C                                                                      C
C     FOR A HARMONIC POTENTIAL V(NOE) = SUM OVER ALL SPECIFIED ATOM    C
C     PAIRS I-J for                                                    C
C     VDIS (NOE) = 0.5*CNOE*(NOE- RNOE0)**2                            C
C                                                                      C
C     PERIODIC BOUNDARY CONDITIONS CANNOT BE TAKEN INTO ACCOUNT        C
C                                                                      C
C     ONLY ONE IDENTICAL MOLECULE CAN BE CONSIDERED.                   C
C                                                                      C
C     NB = TOTAL NUMBER OF NOE RESTRAINTS PER MOLECULE                 C
C                                                                      C
C     IB,JB(1..NB) = ATOMS FORMING NOE RESTRAINT PAIR I-J              C
C     NM = NUMBER OF IDENTICAL MOLECULES WITH NB NOE RESTRAINTS        C
C     NAM = NUMBER OF ATOMS PER MOLECULE (NAM>1)                       C
C     CB = FORCE CONSTANT CNOE                                         C
C     B0(1..NB) = OBS0, SEE ABOVE                                      C
C     BD(1..NB) = NOE       (CALCULATED FROM PRESENT STRUCTURE)        C
C     X(1.. ) = ATOM CARTESIAN COORDINATES; X- AND Y-AXES              C
C     F(1.. ) = DELIVERED WITH THE NOE RESTRAINT FORCES ADDED TO F     C
C     EB = DELIVERED WITH THE TOTAL DISTANCE RESTRAINT ENERGY          C
C                                                                      C
C     (NF=CAN ONLY BE 0 FOR NOERE)                                     C
C     NF  = 0 : TOTAL ENERGY AND FORCES ARE DELIVERED                  C
C     ILDR    INTEGER LIST IN NOE RESTRAINT                            C
C     NILDR   NUMBER IN LIST                                           C
C     PM      PROTON MOTION FACTOR                                     C
C     NPRFL   NUMBER IN PROTON FILE LIST                               C
C     TMIX, TMIX2  MIXING TIMES                                        C
C     NNR1, NNR2   NUMBER OF RESTRAINTS                                C
C     TAUC      CORRELATION TIME                                       C
C     RADIUS    INCLUSION RADIUS                                       C
C     VOMEGA    SPECTROMETER FREQUENCY                                 C
C     NPROT     NUMBER OF PROTONS                                      C
C     DBYD      PROPER DERIVATIVE FORCES                               C
C     NOEDV     NOEDV DERIVATIVE PROCEDURE                             C
C        (0=FAST,Approx. >0=SLOW,Correct)                              C
C     NSTEP     Do full derivative every NSTEP steps                   C
C     NOERPT    0=off 1=on (report calculated and observed NOEs)       C
C                                                                      C
C      EDMONTON, MARCH 6, 1989  PRESENTLY, THE FILE WILL ONLY WORK     C
C      WITH NON-EXCHANGEABLE PROTONS WITH IAC CODE OF '38'             C
C     THIS SUBROUTINE ALLOWS A FULL DERIVATIVE                         C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
      REAL RELAX(150,150), WW0(150,150), WW1(150,150), SPECTD(3)
```

```
      REAL WW2(150,150), XYZS(150,3)
      REAL XYZ(250,3), DIST(250,250), XPEAKM(250,250)
      REAL UPTRI(11325), RESULT(22500), LINV(22500)
      REAL VECTOR(150,150), VECINV(150,150),DISTS(150,150)
      REAL EXDIAR(150), VEC(150)
      REAL TD(150), XPEAK(150), DUM1(150), DUM2(150),BD(650)
      INTEGER NPRSFL(150), NPROTS, MAXPRS, MAXPKS
      REAL PMS(150), TC, TDM(150,150), XPEAKS(150,150)
      REAL XPEAKX(150,150,4),XPEAKI(150),XPEAKJ(150),XPEAK0(150)
      REAL XPEAK2(250,250)
      REAL XPKX2(150,150,4),XPKI2(150),XPKJ2(150),XPK02(150)
C     Ensure that MAXPRS (MAX num of protons in subset)
C     is compatible with array sizes (change MATMLT as well)
C     Ensure that MAXPKS (MAX num of Xpeaks in subset)
C     is compatible with array sizes
C     Check values of NPROTS for MAXPRS and NPROT for 250
C     and MAXPKS > NOXP1+NOXP2
C
      INCLUDE 'em.common.2'
C
      DIMENSION IB(1),JB(1),B0(1),BOX(1),X(1),F(1),XB0(1),
     2          EB0(1),XIJ(3)
C
      MAXPRS=150
      MAXPKS=99
      EB=0.E0
      NN=0
C
C   CALCULATION OF CONSTANTS
C   QCONST=HBAR*GAMMA IN CGS*1E42,  SET UP W MATRIX
      QCONST= 5.671E+04
C    READ IN COORDINATES FOR PROTONS IN NMETERS
C    DISTANCES ARE IN NANOMETERS FOR OUTPUT
      NPROTS=0
C
C        GET COORS USING PROTON FILE
      J=0
      DO 20 I=1,NPROT
      J=NPRFL(I)
      DO 10 M=1,3
      XYZ(I,M)=X(3*(J-1)+M)
   10 CONTINUE
   20 CONTINUE
C    CALCULATE ALL INTERPROTON DISTANCES
      DO 40 I=1,NPROT
      DO 30 J=1,NPROT
      DIST(I,J)=SQRT((XYZ(I,1)-XYZ(J,1))**2+(XYZ(I,2)-XYZ(J,2))**2+
     @(XYZ(I,3)-XYZ(J,3))**2)
   30 CONTINUE
   40 CONTINUE
      DO 45 I=1,NPROT
      DO 42 J=1,NPROT
      XPEAKM(I,J)=0.0
   42 CONTINUE
   45 CONTINUE
C
C
C   DO 'FALSE' BUT QUICK DIRECT NOE DERIVATIVE
```

```
C
C   SET UP LIST OF PROTONS NEAR REFERENCE PROTON (M)
C   AND CALCULATE DISTANCES, XPEAKS TO THAT PROTON
C
      IF(NOERPT .LT. 0) WRITE(6,3000) NSTEP
 3000 FORMAT(' NSTEP IS: ',I5)
      DO 290 M=1,NPROT
C
C     CHECK IF  Proton M IS ON DIS RESTRAINT FILE
C
      I=0
   52 I=I+1
      IF(NPRFL(M).EQ.ILDR(I)) GO TO 54
      IF(I.NE.NILDR) GO TO 52
      GO TO 290
   54 I3=3*(NPRFL(M)-1)
      XYZS(1,1)=XYZ(M,1)
      XYZS(1,2)=XYZ(M,2)
      XYZS(1,3)=XYZ(M,3)
      L=1
      NPRSFL(L)=NPRFL(M)
      PMS(L)=PM(M)
      L=2
      DO 60 J=1,NPROT
      DSCK=DIST(M,J)
      IF (DSCK.GT.RADIUS) GO TO 60
      IF (DSCK.EQ.0.0) GO TO 60
      DO 50 K=1,3
      XYZS(L,K)=XYZ(J,K)
   50 CONTINUE
      NPRSFL(L)=NPRFL(J)
      PMS(L)=PM(J)
      L=L+1
   60 CONTINUE
      NPROTS=L-1
      IF(NPROTS.GT.MAXPRS) THEN
        WRITE(6,*)'NPROTS > MAXPRS (NOEDV>0)', NPROTS, MAXPRS
   STOP
      ENDIF
C    CALCULATE INTERPROTON DISTANCES IN SUBSET
      DO 80 I=1,NPROTS
      DO 70 J=1,NPROTS
      DISTS(I,J)=SQRT((XYZS(I,1)-XYZS(J,1))**2+
     @(XYZS(I,2)-XYZS(J,2))**2+(XYZS(I,3)-XYZS(J,3))**2)
   70 CONTINUE
   80 CONTINUE
C
C
      DO 100 I=1,NPROTS
      DO 90 J=1,NPROTS
      IF (J.EQ.I) GO TO 90
      TC=TAUC*PMS(I)*PMS(J)
      SPECTD(1)=TC
      SPECTD(2)=TC/(1+(VOMEGA*TC)**2)
      SPECTD(3)=TC/(1+(2*VOMEGA*TC)**2)
      WW0(I,J)=QCONST*SPECTD(1)/(DISTS(I,J))**6
      WW1(I,J)=1.5*QCONST*SPECTD(2)/(DISTS(I,J))**6
      WW2(I,J)=6.0*QCONST*SPECTD(3)/(DISTS(I,J))**6
```

```
   90 CONTINUE
  100 CONTINUE
C    SETUP RELAXATION MATRIX
       DO 140 I=1,NPROTS
       DO 130 J=1,NPROTS
       RELAX(I,J)=0.0
       IF (I.NE.J) GO TO 120
       DO 110 K=1,NPROTS
       IF (I.EQ.K) GOTO 110
       RELAX(I,I)=RELAX(I,I)+WW0(I,K)+2*WW1(I,K)+WW2(I,K)
  110 CONTINUE
       GO TO 130
  120 RELAX(I,J)=WW2(I,J)-WW0(I,J)
  130 CONTINUE
  140 CONTINUE
C     SETUP UPPER TRIANGLE COLUMNWISE
       L=1
       DO 160 I=1,NPROTS
       DO 150 J=1,I
       UPTRI(L)=RELAX(I,J)
       L=L+1
  150 CONTINUE
  160 CONTINUE
C    CALCULATE DIAGONAL
       CALL EIGEN(UPTRI,RESULT,NPROTS,0)
C    PUT RESULT INTO MATRIX FORM
       L=1
       DO 180 I=1,NPROTS
       DO 170 J=1,NPROTS
       VECTOR(J,I)=RESULT(L)
       L=L+1
  170 CONTINUE
  180 CONTINUE
C    CALCULATE THE INVERSE FOR VECTOR AND PUT IN VECINV
       L=1
       DO 200 I=1,NPROTS
       DO 190 J=1,NPROTS
       LINV(L)=VECTOR(I,J)
       L=L+1
  190 CONTINUE
  200 CONTINUE
       CALL MINV(LINV,NPROTS,D,DUM1,DUM2)
       L=1
       DO 220 I=1,NPROTS
       DO 210 J=1,NPROTS
       VECINV(I,J)=LINV(L)
       L=L+1
  210 CONTINUE
  220 CONTINUE
C   CALCULATE CROSS PEAKS
       L=1
       DO 230 I=1,NPROTS
       II=I*(I+1)/2
       EXDIAR(L)=EXP(-1*UPTRI(II)*TMIX)
       L=L+1
  230 CONTINUE
       DO 240 K=1,NPROTS
       VEC(K)=VECTOR(1,K)
```

```
  240 CONTINUE
      DO 250 J=1,NPROTS
      DISCK=DISTS(1,J)
      IF (DISCK.EQ.0.0) GO TO 250
      XPEAK(J)=0.0
      DO 245 L=1,NPROTS
      TD(L)=VEC(L)*EXDIAR(L)
      XPEAK(J)=XPEAK(J)+TD(L)*VECINV(L,J)
  245 CONTINUE
      K=0
  260 K=K+1
      IF (NPRFL(K).NE.NPRSFL(J)) GOTO 260
      NPRF=K
      XPEAKM(M,NPRF)=XPEAK(J)
  250 CONTINUE
      IF(NNR1.EQ.NB) GO TO 290
      L=1
      DO 255 I=1,NPROTS
      II=I*(I+1)/2
      EXDIAR(L)=EXP(-1*UPTRI(II)*TMIX2)
      L=L+1
  255 CONTINUE
      DO 280 J=1,NPROTS
      DISCK=DISTS(1,J)
      IF (DISCK.EQ.0.0) GO TO 280
      XPEAK(J)=0.0
      DO 265 L=1,NPROTS
      TD(L)=VEC(L)*EXDIAR(L)
      XPEAK(J)=XPEAK(J)+TD(L)*VECINV(L,J)
  265 CONTINUE
      K=0
  270 K=K+1
      IF (NPRFL(K).NE.NPRSFL(J)) GOTO 270
      NPRF=K
      XPEAK2(M,NPRF)=XPEAK(J)
  280 CONTINUE
  290 CONTINUE
C
C
C
  300 IM=NN*NAM-1
      NNB=NN*NB
      NN=NN+1
      N=0
      RFACTR=0.0
      B0TOT=0.0
      BDTOT=0.0
C
C     CALCULATE NOE SCALING FACTOR
C
  305 N=N+1
      I=0
  340 I=I+1
      IF(NPRFL(I).NE.IB(N)) GOTO 340
      NPRFI=I
      I=0
  370 I=I+1
      IF(NPRFL(I).NE.JB(N)) GOTO 370
```

```
      NPRFJ=I
      BD(N)=XPEAKM(NPRFI,NPRFJ)
      DO 373 K=1,NOM
      IF(NPRFL(NPRFI).EQ.MGRP(K)) BD(N)=BD(N)+XPEAKM(NPRFI+1,NPRFJ)
     @+XPEAKM(NPRFI+2,NPRFJ)
  373 CONTINUE
      DO 375 K=1,NOM
      IF(NPRFL(NPRFJ).EQ.MGRP(K)) BD(N)=BD(N)+XPEAKM(NPRFI,NPRFJ+1)
     @+XPEAKM(NPRFI,NPRFJ+2)
  375 CONTINUE
      IF(BD(N).EQ.0.0) WRITE(6,3004)
      IF(DIST(NPRFI,NPRFJ).LT.RADIUS) GOTO 1002
      WRITE(6,3003) N,NPRFI,NPRFJ,DIST(NPRFI,NPRFJ),RADIUS
      WRITE(6,*) 'IB, JB,',IB(N),JB(N),NPRFL(NPRFI),NPRFL(NPRFJ)
 3003 FORMAT('DIST GT RADIUS N,I,J,DIST, &RADIUS ARE:',3I4,2F6.3)
 3004 FORMAT(' CALC OR STORED NOE IS EQUAL TO ZERO')
 1002 BOTOT=BOTOT+BO(N)
      BDTOT=BDTOT+BD(N)
      IF(N.LT.NNR1) GOTO 305
      IF(NB.EQ.NNR1) GOTO 345
  325 N=N+1
      I=0
  330 I=I+1
      IF(NPRFL(I).NE.IB(N)) GOTO 330
      NPRFI=I
      I=0
  335 I=I+1
      IF(NPRFL(I).NE.JB(N)) GOTO 335
      NPRFJ=I
      BD(N)=XPEAK2(NPRFI,NPRFJ)
      DO 383 K=1,NOM
      IF(NPRFL(NPRFI).EQ.MGRP(K)) BD(N)=BD(N)+XPEAKM(NPRFI+1,NPRFJ)
     @+XPEAKM(NPRFI+2,NPRFJ)
  383 CONTINUE
      DO 385 K=1,NOM
      IF(NPRFL(NPRFJ).EQ.MGRP(K)) BD(N)=BD(N)+XPEAKM(NPRFI,NPRFJ+1)
     @+XPEAKM(NPRFI,NPRFJ+2)
  385 CONTINUE
      IF(BD(N).EQ.0.0) WRITE(6,3004)
      IF(DIST(NPRFI,NPRFJ).LT.RADIUS) GOTO 1052
      WRITE(6,3003) N,NPRFI,NPRFJ,DIST(NPRFI,NPRFJ),RADIUS
      WRITE(6,*) 'IB, JB,',IB(N),JB(N),NPRFL(NPRFI),NPRFL(NPRFJ)
 1052 BOTOT=BOTOT+BO(N)
      BDTOT=BDTOT+BD(N)
      IF(N.LT.NB) GOTO 325
  345 SFNOE=BDTOT/BOTOT
      N=0
  310 N=N+1
      I3=3*(IB(N)+IM)
      DO 313 K=1,NOM
      IF(IB(N).EQ.MGRP(K)) I3=3*(MGRPC(K)-1)
  313 CONTINUE
      J3=3*(JB(N)+IM)
      DO 315 K=1,NOM
      IF(JB(N).EQ.MGRP(K)) J3=3*(MGRPC(K)-1)
  315 CONTINUE
      RNOE0=BO(N)*SFNOE
      RNOE=BD(N)
```

```
      RIJ2=0.E0
      DO 380 M=1,3
      XIJ(M)=X(I3+M)-X(J3+M)
  380 RIJ2=RIJ2+XIJ(M)**2
      RIJ= SQRT(RIJ2)
      DBF=(1.0/(RNOE**0.167))-(1.0/(RNOE0**0.167))
      DB=RNOE-RNOE0
C
C ADDING NOERPT June 30 89
      IF (NOERPT .NE. 1) GOTO 381
      WRITE(6,3005) N,IB(N),JB(N),RNOE0,RNOE,DB
 3005 FORMAT(' N,I,J,OBS,CALC,DIFF ARE:',3I4,3F7.4)
  381 DF=0.E0
      EBH=0.E0
      DF=CB*DBF
      EBH=CB*DB*DB
      RFACTR=RFACTR+ABS(DB)
      EB=EB+EBH
C
      IF(NOEDV.NE.0) GO TO 441
      DF=DF/RIJ
      DO 440 M=1,3
      XH=XIJ(M)*DF
      F(I3+M)=F(I3+M)-XH
      F(J3+M)=F(J3+M)+XH
  440 CONTINUE
  441 IF (N.LT.NB) GOTO 310
      RFACTR=RFACTR/(B0TOT*SFNOE)
      WRITE(6,1023) SFNOE,RFACTR
 1023 FORMAT(1X,'NOE SCALING AND R FACTOR ARE:',2F8.4)
      IF (NN.LT.NM) GOTO 300
      EB=EB/2.E0
      IF(NOEDV.EQ.0) GO TO 900
C
C     NOW DO PROPER DERIVATIVE AT THE ZERO AND EVERY NOEDV STEPS
C
      IF(NSTEP.NE.NSTEP/NOEDV*NOEDV) GO TO 900
      DO 890 M=1,NPROT
      IF(NOERPT .LT. 0) WRITE(6,3001) NSTEP,M
 3001 FORMAT('NSTEP, M are :',2I5)
C
      XYZS(1,1)=XYZ(M,1)
      XYZS(1,2)=XYZ(M,2)
      XYZS(1,3)=XYZ(M,3)
      L=1
      NPRSFL(L)=NPRFL(M)
      PMS(L)=PM(M)
      L=2
      DO 460 J=1,NPROT
      DSCK=DIST(M,J)
      IF (DSCK.GT.RADIUS) GO TO 460
      IF (DSCK.EQ.0.0) GO TO 460
      DO 450 K=1,3
      XYZS(L,K)=XYZ(J,K)
  450 CONTINUE
      NPRSFL(L)=NPRFL(J)
      PMS(L)=PM(J)
      L=L+1
```

```
  460 CONTINUE
      NPROTS=L-1
      IF(NPROTS.GT.MAXPRS) THEN
        WRITE(6,*)'NPROTS > MAXPRS (NOEDV>0)', NPROTS, MAXPRS
     STOP
      ENDIF
      NOXP=0
      NOXP2=0
      DO 502 I=1,NB
      DO 501 K=1,NPROTS
      IF(NPRSFL(K).EQ.IB(I))GO TO 503
  501 CONTINUE
      GO TO 502
  503 CONTINUE
      DO 504 L=1,NPROTS
      IF(NPRSFL(L).EQ.JB(I))GO TO 505
  504 CONTINUE
      GO TO 502
  505 IF (I.GT.NNR1) GOTO 1505
      NOXP=NOXP+1
      XPEAKI(NOXP)=K
      XPEAKJ(NOXP)=L
      XPEAK0(NOXP)=B0(I)*SFNOE
      GOTO 502
 1505 NOXP2=NOXP2+1
      XPKI2(NOXP2)=K
      XPKJ2(NOXP2)=L
      XPK02(NOXP2)=B0(I)*SFNOE
  502 CONTINUE
      IF(NOXP.EQ.0.AND.NOXP2.EQ.0) GOTO 890
      IF(NOXP.GT.MAXPKS.OR.NOXP2.GT.MAXPKS) THEN
        WRITE(6,*)'NOXP or NOXP2 > MAXPKS'
        WRITE(6,3002)NSTEP,NOXP,NOXP2,M
        STOP
 3002 FORMAT(' NSTEP, noxp, noxp2, m are:',4I5)
      ENDIF
C
C     SET UP COORD FILE FOR DERIVATIVE, CORRECT IF NECESSARY
      NDV=4
      IF(NDV.EQ.4) GO TO 568
  565 XYZS(1,1)=XYZ(M,1)
      XYZS(1,2)=XYZ(M,2)
      XYZS(1,3)=XYZ(M,3)
      XYZS(1,NDV)=XYZ(M,NDV)+0.0010
  568 CONTINUE
C    CALCULATE INTERPROTON DISTANCES IN SUBSET
      DO 580 I=1,NPROTS
      DO 570 J=1,NPROTS
      DISTS(I,J)=SQRT((XYZS(I,1)-XYZS(J,1))**2+
     @(XYZS(I,2)-XYZS(J,2))**2+(XYZS(I,3)-XYZS(J,3))**2)
  570 CONTINUE
  580 CONTINUE
C
C
      DO 600 I=1,NPROTS
      DO 590 J=1,NPROTS
      IF (J.EQ.I) GO TO 590
      TC=TAUC*PMS(I)*PMS(J)
```

```
      SPECTD(1)=TC
      SPECTD(2)=TC/(1+(VOMEGA*TC)**2)
      SPECTD(3)=TC/(1+(2*VOMEGA*TC)**2)
      WW0(I,J)=QCONST*SPECTD(1)/(DISTS(I,J))**6
      WW1(I,J)=1.5*QCONST*SPECTD(2)/(DISTS(I,J))**6
      WW2(I,J)=6.0*QCONST*SPECTD(3)/(DISTS(I,J))**6
  590 CONTINUE
  600 CONTINUE
C   SETUP RELAXATION MATRIX
      DO 640 I=1,NPROTS
      DO 630 J=1,NPROTS
      RELAX(I,J)=0.0
      IF (I.NE.J) GO TO 620
      DO 610 K=1,NPROTS
      IF (I.EQ.K) GOTO 610
      RELAX(I,I)=RELAX(I,I)+WW0(I,K)+2*WW1(I,K)+WW2(I,K)
  610 CONTINUE
      GO TO 630
  620 RELAX(I,J)=WW2(I,J)-WW0(I,J)
  630 CONTINUE
  640 CONTINUE
C    SETUP UPPER TRIANGLE COLUMNWISE
      L=1
      DO 660 I=1,NPROTS
      DO 650 J=1,I
      UPTRI(L)=RELAX(I,J)
      L=L+1
  650 CONTINUE
  660 CONTINUE
C   CALCULATE DIAGONAL
      CALL EIGEN(UPTRI,RESULT,NPROTS,0)
C   PUT RESULT INTO MATRIX FORM
      L=1
      DO 680 I=1,NPROTS
      DO 670 J=1,NPROTS
      VECTOR(J,I)=RESULT(L)
      L=L+1
  670 CONTINUE
  680 CONTINUE
C   CALCULATE THE INVERSE FOR VECTOR AND PUT IN VECINV
      L=1
      DO 700 I=1,NPROTS
      DO 690 J=1,NPROTS
      LINV(L)=VECTOR(I,J)
      L=L+1
  690 CONTINUE
  700 CONTINUE
      CALL MINV(LINV,NPROTS,D,DUM1,DUM2)
      L=1
      DO 720 I=1,NPROTS
      DO 710 J=1,NPROTS
      VECINV(I,J)=LINV(L)
      L=L+1
  710 CONTINUE
  720 CONTINUE
C  CALCULATE CROSS PEAKS
      L=1
      DO 730 I=1,NPROTS
```

```
      II=I*(I+1)/2
      EXDIAR(L)=EXP(-1*UPTRI(II)*TMIX)
      L=L+1
  730 CONTINUE
      DO 764 I=1,NPROTS
      DO 762 J=1,NPROTS
      TDM(I,J)=VECTOR(I,J)*EXDIAR(J)
  762 CONTINUE
  764 CONTINUE
      CALL MATMLT(TDM,VECINV,XPEAKS,NPROTS)
      DO 785 I=1,NPROTS
      DO 780 J=1,NPROTS
      IF (J.EQ.I) GO TO 780
      XPEAKX(I,J,NDV)=XPEAKS(I,J)
  780 CONTINUE
  785 CONTINUE
C
C Calculate NOE for second mixing time.
C
      IF(NNR1.EQ.NB) GO TO 1786
      L=1
      DO 1730 I=1,NPROTS
      II=I*(I+1)/2
      EXDIAR(L)=EXP(-1*UPTRI(II)*TMIX2)
      L=L+1
 1730 CONTINUE
      DO 1764 I=1,NPROTS
      DO 1762 J=1,NPROTS
      TDM(I,J)=VECTOR(I,J)*EXDIAR(J)
 1762 CONTINUE
 1764 CONTINUE
      CALL MATMLT(TDM,VECINV,XPEAKS,NPROTS)
      DO 1785 I=1,NPROTS
      DO 1780 J=1,NPROTS
      IF (J.EQ.I) GO TO 1780
      XPKX2(I,J,NDV)=XPEAKS(I,J)
 1780 CONTINUE
 1785 CONTINUE
 1786 NDV=NDV-1
      IF(NDV.NE.0) GO TO 565
C
C     SUM PEAKS, CALCULATE DERIVATIVE
      I3=3*(NPRFL(M)-1)
      DO 802 L=1,3
      DBYD(I3+L)=0.0
      DO 801 K=1,NOXP
      I=XPEAKI(K)
      J=XPEAKJ(K)
      DBYD(I3+L)=DBYD(I3+L)+CB*1000*(XPEAKX(I,J,4)-XPEAK0(K))
     @*(XPEAKX(I,J,4)-XPEAKX(I,J,L))
  801 CONTINUE
      IF(NNR1.EQ.NB) GO TO 802
C
C Calculating derivative for second set of Noes.
C
      DO 1801 K=1,NOXP2
      I=XPKI2(K)
      J=XPKJ2(K)
```

```
      DBYD(I3+L)=DBYD(I3+L)+CB*1000*(XPKX2(I,J,4)-XPK02(K))
     @*(XPKX2(I,J,4)-XPKX2(I,J,L))
 1801 CONTINUE
  802 CONTINUE
  890 CONTINUE
C
C     CALC FORCES
  900 CONTINUE
      DO 940 I=1,NPROT
      I3=3*(NPRFL(I)-1)
      DO 930 K=1,NOM
      IF(NPRFL(I).EQ.MGRP(K)) I3=3*(MGRPC(K)-1)
  930 CONTINUE
      DO 920 J=1,3
      F(I3+J)=F(I3+J)+DBYD(I3+J)
  920 CONTINUE
  940 CONTINUE
      RETURN
      END
      SUBROUTINE MATMLT(VMAT1,VMAT2,RMAT,N)
C
      REAL VMAT1(150,150), VMAT2(150,150), RMAT(150,150), TOT
      DO 3 I=1,N
    DO 2 J=1,N
            TOT = 0.0
      DO 1 K=1,N
       TOT=TOT+VMAT1(I,K)*VMAT2(K,J)
    1    CONTINUE
      RMAT(I,J)=TOT
    2  CONTINUE
    3  CONTINUE
      RETURN
      END
```

Appendix 2 Interproton distances in right-handed helices

Nuclei		Distance to 2nd nucleus (Å) A DNA			B DNA		
1st	2nd	i-1	i	i+1	i-1	i	i+1
XHB	XHB	4.7	na	4.7	5.0	na	5.0
XHB	XH1'	4.1	3.8	>	2.8	3.8	>
XHB	XH2'	1.6	3.7	>	3.8	2.0	>
XHB	XH2''	3.2	4.6	>	2.3	3.6	>
XHB	XH3'	3.4	2.7	>	5.0	4.0	>
XHB	XH4'	>	4.1	>	>	4.6	>
XHB	XH5'	>	3.4	>	>	3.4	>
XHB	XH5''	>	4.1	>	>	4.2	>
XHB	TCH3	>	2.7	3.1	>	2.7	3.1
XHB	CH5	>	2.5	3.8	>	2.4	3.8
AH2	XH1'	>	4.6	4.0	>	4.5	4.6
AH2	XHB	>	>	4.7	>	>	>
AH2	XH2'	>	5.0	4.7	>	>	>
AH2	CH5	>	>	5.0	>	>	>
AH2	AH2	4.9	na	4.9	3.8	na	3.8
CH5	CH5	3.8	na	3.8	4.6	na	4.6
CH5	XH1'	4.8	>	>	3.7	>	>
CH5	XH2''	4.8	>	>	2.8	>	>
CH5	XH2'	3.0	>	>	3.3	4.2	>
CH5	XH3'	3.6	4.5	>	>	>	>
CH5	TCH3	4.3	na	2.9	>	>	3.9
XH1'	XH1'	>	na	>	4.9	na	4.9
XH1'	XH2''	>	2.3	>	>	2.3	>
XH1'	XH2'	4.6	2.7	>	>	3.0	4.1
XH1'	XH3'	>	3.8	>	>	3.9	4.9
XH1'	XH4'	>	3.3	>	>	3.6	4.2
XH1'	XH5'	>	4.6	3.8	>	4.5	1.8
XH1'	XH5''	>	>	>	>	>	3.3

XH1'	TCH3	>	>	4.7	>	>	3.9
XH2"	XH2'	>	1.8	>	>	1.8	3.8
XH2"	XH3'	>	3.0	4.6	>	2.5	5.0
XH2"	XH4'	>	2.8	4.1	>	4.1	>
XH2"	XH5'	>	4.9	1.7	>	4.9	3.4
XH2"	XH5''	>	4.7	3.3	>	4.9	4.1
XH2"	TCH3	>	>	4.6	>	>	2.3
XH2'	XH3'	>	2.4	3.7	>	2.4	>
XH2'	XH4'	>	3.8	4.3	>	3.8	>
XH2'	XH5'	>	>	2.6	>	3.8	4.5
XH2'	XH5''	>	>	3.8	>	3.9	>
XH2'	TCH3	>	>	2.8	>	4.5	2.8
XH3'	XH4'	>	3.0	>	>	2.7	>
XH3'	XH5'	>	3.6	4.2	>	3.7	4.6
XH3'	XH5''	>	3.1	4.9	>	2.8	4.8
XH3'	TCH3	>	4.9	2.7	>	>	4.8
XH4'	XH5'	>	2.5	4.1	>	2.6	3.8
XH4'	XH5''	>	2.4	>	>	2.3	4.2
XH5'	TCH3	>	>	>	>	>	>
		j-1	j	j+1	j-1	j	j+1
AH2	AH2	>	>	4.9	4.1	>	3.6
AH2	XH1'	>	>	4.7	>	>	>
		3.9	>	>			

Protons were placed in geometrically reasonable positions assuming a C-H bond length of 1.08 Å on the A and B forms of DNA (Arnott & Hukins, 1972). Because of small differences between distances associated with purine and pyrimidines, these distances are accurate to approximately 0.2 Å. All distances are given in Angstroms (= 0.1 nm.). The nomenclature for designating protons is that used in any standard text. "B" refers to the H6 or H8 base proton. A distance to a thymine methyl

(M) is shown as r6 weighted average. Data is included here for non-exchangeable protons only. A > sign indicates that the distance is greater than 5 Å.

Nucleotides are given an i (or j) +/- 1 designation in a 5'-3' orientation:

5' (i - 1) - (i) - (i + 1) 3'
3' (j + 1) - (i) - (j - 1) 5'

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* 47,1504-1509.

Appendix 3 Experimental restraints and coordinates for GTAC[a]

A. NOE intensities and distances

Mixing time (sec.)				Distance estimate (nm.)		Proton pair
0.100	0.200	0.500	0.500	Lower	Upper	
1. Resolved cross-peaks						
0.408	0.830	0.939	0.911	0.270	0.310	1dG 1' - 1dG 2'
0.936	1.579	1.363	1.553	0.230		1dG 1' - 1dG 2"
0.081	0.299	0.403	0.464	0.350	0.420	1dG 1' - 2dT H6
	0.123	0.489	0.393	0.400	0.550	1dG 1' - 2dT HM
			0.194			1dG 2" - 2dT 2'
0.469	0.767	0.690	0.619	0.260	0.290	1dG 2" - 2dT H6
	0.149	0.887	0.684	0.400	0.560	1dG 2" - 2dT HM
0.573	0.060[a]	1.200		0.230	0.260	1dG 2' - 1dG 3'
0.179	0.404	0.802	0.564	0.320	0.380	1dG 4' - 1dG 1'
0.000		0.063	0.149			1dG 4' - 1dG H8
	0.224	0.436	0.472	0.360	0.420	1dG H8 - 1dG 1'
0.093	0.260	0.540	0.501	0.350	0.430	1dG H8 - 1dG 2"
0.513	1.076	1.164	1.049	0.250	0.290	1dG H8 - 1dG 2'
0.309	0.576	0.876	0.836	0.330	0.410	1dG H8 - 2dT HM
1.612	2.160	2.375	2.249	0.230		2dT 1' - 2dT 2"
0.043	0.376	0.871	0.729	0.370	0.440	2dT 1' - 2dT H6
0.052	0.182	0.471	0.424	0.390	0.440	2dT 1' - 3dA H8
0.455	1.111	1.464	1.445	0.260	0.360	2dT 2" - 2dT H6
0.365	0.702	0.735	0.642	0.270	0.320	2dT 2" - 3dA H8
0.614	1.013	1.653	1.435	0.260	0.285	2dT 2' - 2dT 1'
1.489	1.844	1.752	1.747	0.220	0.240	2dT 2' - 2dT 3'
2.489	2.924	2.240	1.712	0.200	0.220	2dT 2' - 2dT H6
			0.355			2dT 2' - 2dT HM
0.125	0.346	0.346	0.355	0.310	0.380	2dT 2' - 3dA H8
0.000	0.085	0.503	0.750			2dT 3' - 2dT 1'
0.057	0.525	1.146	0.839	0.350	0.500	2dT 3' - 2dT H6
0.076		0.238	0.248	0.340	0.440	2dT 3' - 3dA H8
0.211	0.671	1.038	1.050	0.290	0.360	2dT 4' - 2dT 1'
		0.569	0.360			2dT 4' - 2dT 2"
		0.475	0.270			2dT 4' - 2dT 2'
0.809	1.187	1.947	1.742	0.250	0.340	2dT 4' - 2dT 3'
0.000	0.065	0.382	0.331	0.400	0.560	2dT 4' - 2dT H6
0.318	0.570	0.577	0.439	0.280	0.360	2dT H6 - 1dG 2'
0.000		0.159	0.182			2dT H6 - 3dA H8
	0.326	1.064	0.857	0.370	0.500	2dT HM - 1dG 2'
	0.160	0.795	0.723	0.370	0.500	3dA 1' - 3dA H8
0.000		0.239	0.145			3dA 1' - 4dC 1'
	0.123	0.335	0.365	0.380	0.480	3dA 1' - 4dC H5
0.078	0.346	0.722	0.594	0.340	0.480	3dA 1' - 4dC H6
0.426	0.900	1.422	1.450	0.270	0.360	3dA 2" - 3dA H8
	0.354	0.690	0.590	0.280	0.400	3dA 2" - 4dC H5
0.690	1.020	0.801	0.860	0.240	0.270	3dA 2" - 4dC H6
1.715	2.000	2.199	1.798	0.210	0.240	3dA 2' - 3dA 3'
2.142	2.400	2.644	1.954	0.200	0.230	3dA 2' - 3dA H8

	0.203	0.574	0.414	0.330	0.450	3dA 2' - 4dC H5
1.650	1.400	1.680	1.448	0.210	0.270	3dA 3' - 3dA 2"
	0.365	0.865	0.909	0.310	0.420	3dA 3' - 3dA H8
0.000	0.029	0.276	0.361			3dA 3' - 4dC H6
0.170	0.360	0.900	0.765	0.330	0.380	3dA 4' - 3dA 2"
0.084	0.230	0.700	0.438	0.340	0.410	3dA 4' - 3dA 2'
0.980	1.700	2.300	1.569	0.230	0.270	3dA 4' - 3dA 3'
0.000		0.262	0.268			3dA 4' - 3dA H8
0.026		0.208		0.360	0.540	3dA H2 - 4dC 1'
0.057	0.270	0.500	0.429	0.350	0.500	3dA H8 - 4dC H5
0.000	0.075	0.316	0.363	0.420	0.570	3dA H8 - 4dC H6
1.534	2.060	1.996	2.189	0.230		4dC 1' - 4dC 2"
0.046	0.224	0.599	0.531	0.390	0.450	4dC 1' - 4dC H6
0.138	0.369	0.530	0.665	0.330	0.390	4dC 1' - 5dG H8
		0.295	0.247			4dC 2" - 4dC H5
0.319	0.799	0.808	0.977	0.280	0.380	4dC 2" - 4dC H6
0.372	0.635	0.538	0.577	0.270	0.300	4dC 2" - 5dG H8
0.461	0.740	1.235	1.290	0.270	0.300	4dC 2' - 4dC 1'
0.662	0.950[a]	0.553		0.210	0.250	4dC 2' - 4dC 3'
	0.227	0.588	0.458	0.330	0.460	4dC 2' - 4dC H5
0.167	0.295	0.317	0.305	0.320	0.370	4dC 2' - 5dG H8
0.043	0.440[b]	0.260		0.330	0.440	4dC 3' - 4dC H6
0.313	0.818	1.602	1.696	0.280	0.360	4dC 4' - 4dC 1'
1.410	2.141	2.074	2.024	0.230		5dG 1' - 5dG 2"
0.115	0.228	0.727	0.665	0.370	0.410	5dG 1' - 5dG H8
0.149	0.294	0.676	0.815	0.350	0.380	5dG 1' - 6dT H6
0.075	0.234	0.687	0.500	0.440	0.560	5dG 1' - 6dT HM
0.221	0.911	1.114	0.769	0.280	0.370	5dG 2" - 5dG H8
0.000			0.200			5dG 2" - 6dT 2'
	0.267	1.335	0.742	0.360	0.500	5dG 2" - 6dT HM
0.531	1.163	1.568	1.412	0.260	0.300	5dG 2' - 5dG 1'
0.199	0.506[a]	0.830		0.210	0.250	5dG 2' - 5dG 3'
1.524	2.030	1.618	1.422	0.215	0.235	5dG 2' - 5dG H8
0.247	0.521	0.864	0.567	0.300	0.370	5dG 2' - 6dT H6
	0.295	1.284	0.807	0.400	0.490	5dG 2' - 6dT HM
0.058	0.131	0.919	0.636	0.360	0.500	5dG 3' - 5dG 1'
0.801	1.360	1.692	1.765	0.250	0.270	5dG 3' - 5dG 2"
	0.317	0.671	0.596	0.320	0.400	5dG 3' - 5dG H8
		0.330	0.271			5dG 3' - 6dT H6
0.055	0.095	0.330	0.271	0.380	0.470	5dG 3' - 6dT H6
0.212	0.689	0.752	0.935	0.290	0.360	5dG 4' - 5dG 1'
	0.092	0.424	0.308	0.380	0.500	5dG 4' - 5dG 2"
	0.145	0.278	0.336	0.360	0.450	5dG 4' - 5dG 2'
1.011	1.615	1.921	1.946	0.240	0.260	5dG 4' - 5dG 3'
0.000	0.058	0.226	0.200	0.400	0.550	5dG 4' - 5dG H8
0.615	0.890	1.308	1.328	0.310	0.350	5dG H8 - 6dT HM
1.704	2.568	2.184	2.781	0.230		6dT 1' - 6dT 2"
0.143	0.312	0.759	0.793	0.350	0.380	6dT 1' - 6dT H6
0.041	0.259	0.419	0.489	0.380	0.470	6dT 1' - 7dA H8
0.422	1.043	1.500	1.389	0.270	0.360	6dT 2" - 6dT H6
0.355	0.750	0.640	0.507	0.260	0.300	6dT 2" - 7dA H8
0.469	0.942	1.399	1.502	0.270	0.320	6dT 2' - 6dT 1'
1.180	1.127[a]	2.039	1.845	0.210	0.260	6dT 2' - 6dT 3'
0.000		0.089	0.202			6dT 2' - 6dT HM
0.262	0.483	0.696	0.499	0.290	0.360	6dT 2' - 7dA H8
		0.590	0.748			6dT 3' - 6dT 1'
0.745	1.016	1.617	1.394	0.250	0.280	6dT 3' - 6dT 2"
	0.440	0.848	1.066	0.290	0.370	6dT 3' - 6dT H6

		0.239	0.171			6dT 3' - 7dA H8
0.000		0.261	0.197			6dT H6 - 7dA H8
0.032	0.377	0.795	0.707	0.330	0.460	7dA 1' - 7dA H8
		0.397	0.155			7dA 1' - 8dC H5
0.108	0.293	0.798	0.758	0.360	0.410	7dA 1' - 8dC H6
0.399	1.100	1.424	1.417	0.280	0.380	7dA 2" - 7dA H8
	0.249	0.690	0.448	0.320	0.400	7dA 2" - 8dC H5
1.002	1.309	0.977	1.006	0.230	0.250	7dA 2" - 8dC H6
1.715	2.500	2.121	1.621	0.210	0.240	7dA 2' - 7dA 3'
1.970	2.600	2.044	1.942	0.200	0.220[e]	7dA 2' - 7dA H8
	0.168	0.410	0.303	0.320	0.430[e]	7dA 2' - 8dC H5
0.323	0.555	0.782	0.727	0.290	0.350[e]	7dA 2' - 8dC H6
1.650	1.400	1.680	1.742	0.220	0.260	7dA 3' - 7dA 2"
0.083	0.425	0.907	0.995	0.350	0.500	7dA 3' - 7dA H8
0.152	0.103	0.275	0.412	0.320	0.440	7dA 3' - 8dC H6
0.170	0.250	0.529	0.423	0.320	0.380	7dA 4' - 7dA 2"
0.160	0.230	0.400	0.363	0.310	0.380	7dA 4' - 7dA 2'
0.645	1.600	2.300	2.158	0.250	0.290	7dA 4' - 7dA 3'
	0.056	0.265	0.221	0.410	0.560	7dA 4' - 7dA H8
0.000	0.044	0.081	0.156	0.410	0.560	7dA H2 - 8dC 1'
0.114	0.270	0.500	0.493	0.360	0.420	7dA H8 - 8dC H5
0.000	0.064	0.365	0.239	0.450	0.580	7dA H8 - 8dC H6
0.081	0.395	0.805	0.687	0.370	0.450	8dC 1' - 8dC H6
	0.216	0.757	0.727	0.340	0.440	8dC 3' - 8dC 1'
0.274	0.535	0.983	1.068	0.300	0.340	8dC 3' - 8dC H6
0.515	1.294	1.711	1.729	0.260	0.310	8dC 4' - 8dC 1'
	0.950	2.128	2.179	0.260	0.310	8dC 4' - 8dC 3'
0.000	0.142	0.604	0.561	0.340	0.470	8dC 4' - 8dC H6

2. Overlapping cross-peaks

0.054	0.284[b]	0.636	0.548[b]	0.410	0.500	1dG H8 - 1dG 5's
0.101	0.490[b]	1.063	0.740[b]	0.250	0.450	1dG 1' - 2dT 5's
1.848	2.850[b]	3.760	2.848[b]	0.230	0.270	1dG 4' - 1dG 5's
1.579	2.452	2.400	2.590	0.230		3dA 1' - 3dA 2"
						7dA 1' - 7dA 2"[c]
0.624	1.306	1.640	1.737	0.240	0.300	3dA 2' - 3dA 1'
						7dA 2' - 7dA 1'[c]
0.700	0.800			0.210	0.340[e]	3dA 2' - 4dC H6
0.332	0.784	1.400		0.290	0.340[e]	5dG 2" - 6dT H6
0.079	0.300	1.019	0.940	0.330	0.460	3dA 3' - 3dA 1'
0.079	0.383	1.019	0.806	0.330	0.460	7dA 3' - 7dA 1'
0.471	0.727	1.186	1.172	0.260	0.320	3dA 4' - 3dA 1'
0.471	0.727	1.186	1.045	0.260	0.320	7dA 4' - 7dA 1'
1.970	2.316			0.200	0.230[e]	6dT 2' - 6dT H6
						4dC 2' - 4dC H6[d]
0.753	1.305	1.645		0.240	0.290	8dC 3' - 8dC 2's
0.672		1.298				8dC H6 - 8dC 2's
1.441	2.028	2.693		0.230	0.260	8dC 1' - 8dC 2's

3. Covalently bound proton pairs

4.053	4.840	2.615	2.221	0.176		2dT 2' - 2dT 2"
1.086	1.755	2.286	1.967	0.300		2dT H6 - 2dT HM
2.056	3.500	2.007	2.141	0.176		3dA 2' - 3dA 2"
3.500	4.380	2.186	2.491	0.176		4dC 2' - 4dC 2"

1.305	2.155	2.615	2.692	0.246	4dC H5 - 4dC H6
			2.057[a]		5dG 2' - 5dG 2"
3.144	4.380	2.187	2.271	0.176	6dT 2' - 6dT 2"
1.173	1.951	2.433	2.168	0.300	6dT H6 - 6dT HM
2.338	3.500	2.007	2.175	0.176	7dA 2' - 7dA 2"
1.385	2.342	3.110	3.920	0.246	8dC H5 - 8dC H6

[a]The absence of an entry indicates that no distance estimate was made, or that the NOE intensity was not tabulated. 5' and 5" NOE intensities and distances were not used stereospecifically because the ratio of $J_{5'5"}$ to the difference in chemical shifts is often not small, leading to second order effects. 5's and 2's refer to non-stereospecifc NOEs to the protons on the 2' and 5' methylene carbons, respectively. NOE intensities are given for one of the two symmetry related strands. [b]NOE noted in subsequent analysis of the data, but not used in structure refinement. [c]NOE and distances are included with the previous entry. [d]NOE and distances are approximately equal to the previous entry. [e]Distance noted in subsequent analysis and was only used during energy minimization of structures produced from molecular dynamics calculations.

B. Glycosidic dihedral angle restraints

	Allowed ranges of dihedral angles									
	ν_0		ν_1		ν_2		ν_3		ν_4	
Residue										
1	-37	-27	04	39	-30	20	-36	05	20	40
2	-32	-19	33	40	-39	-14	-11	28	-06	32
3	-38	-19	23	40	-39	00	-23	28	-06	38
4	-35	-19	29	40	-39	-07	-18	28	-06	35
5	-38	-19	23	40	-39	00	-23	28	-06	38
6	-32	-19	33	40	-39	-14	-11	28	-06	32
7	-38	-19	23	40	-39	00	-23	28	-06	38
8	-27	-19	04	40	-39	20	-36	28	-06	40

C. Coordinates (in nm.) for $[d(G\text{-}T\text{-}A\text{-}C\text{-}G\text{-}T\text{-}A\text{-}C)]_2$
322 atoms, protons omitted

Atom	No.	x	y	z	Atom	No.	x	y	z
1dG O5*	1	0.522	0.732	1.255	9dG O5*	162	0.226	0.852	-1.244
1dG C5*	2	0.659	0.715	1.290	9dG C5*	163	0.124	0.945	-1.281
1dG C4*	3	0.733	0.628	1.190	9dG C4*	164	0.010	0.945	-1.180
1dG O4*	4	0.673	0.498	1.191	9dG O4*	165	-0.053	0.815	-1.181
1dG C1*	5	0.659	0.454	1.055	9dG C1*	166	-0.082	0.780	-1.045
1dG N9	6	0.544	0.360	1.050	9dG N9	167	-0.088	0.632	-1.036
1dG C4	7	0.556	0.231	1.022	9dG C4	168	-0.198	0.564	-1.007
1dG N3	8	0.660	0.160	0.976	9dG N3	169	-0.317	0.602	-0.958
1dG C2	9	0.652	0.028	0.962	9dG C2	170	-0.416	0.515	-0.941
1dG N2	10	0.759	-0.040	0.924	9dG N2	171	-0.533	0.556	-0.894
1dG N1	11	0.534	-0.039	0.995	9dG N1	172	-0.398	0.380	-0.975
1dG C6	12	0.424	0.033	1.042	9dG C6	173	-0.276	0.338	-1.026
1dG O6	13	0.321	-0.025	1.076	9dG O6	174	-0.262	0.223	-1.067
1dG C5	14	0.436	0.170	1.056	9dG C5	175	-0.175	0.431	-1.041
1dG N7	15	0.354	0.264	1.101	9dG N7	176	-0.051	0.423	-1.087
1dG C8	16	0.421	0.380	1.098	9dG C8	177	0.001	0.545	-1.084

1dG	C2*	17	0.635	0.581	0.975	9dG	C2*	178	0.031	0.841	-0.964
1dG	C3*	18	0.722	0.682	1.048	9dG	C3*	179	0.058	0.971	-1.038
1dG	O3*	19	0.854	0.697	0.992	9dG	O3*	180	-0.014	1.083	-0.984
2dT	P	20	0.881	0.755	0.844	10dT	P	181	0.013	1.142	-0.836
2dT	O1P	21	0.769	0.844	0.804	10dT	O1P	182	0.151	1.109	-0.795
2dT	O2P	22	1.019	0.807	0.839	10dT	O2P	183	-0.032	1.283	-0.833
2dT	O5*	23	0.871	0.619	0.758	10dT	O5*	184	-0.089	1.051	-0.750
2dT	C5*	24	0.969	0.516	0.775	10dT	C5*	185	-0.230	1.068	-0.764
2dT	C4*	25	0.936	0.395	0.691	10dT	C4*	186	-0.306	0.969	-0.679
2dT	O4*	26	0.815	0.329	0.734	10dT	O4*	187	-0.284	0.833	-0.721
2dT	C1*	27	0.757	0.274	0.615	10dT	C1*	188	-0.289	0.752	-0.601
2dT	N1	28	0.613	0.246	0.638	10dT	N1	189	-0.218	0.624	-0.624
2dT	C6	29	0.524	0.348	0.675	10dT	C6	190	-0.082	0.620	-0.660
2dT	C2	30	0.568	0.113	0.643	10dT	C2	191	-0.292	0.505	-0.631
2dT	O2	31	0.636	0.021	0.599	10dT	O2	192	-0.405	0.497	-0.583
2dT	N3	32	0.440	0.085	0.688	10dT	N3	193	-0.232	0.389	-0.679
2dT	C4	33	0.353	0.187	0.727	10dT	C4	194	-0.098	0.387	-0.717
2dT	O4	34	0.242	0.156	0.771	10dT	O4	195	-0.053	0.285	-0.769
2dT	C5	35	0.395	0.319	0.719	10dT	C5	196	-0.022	0.504	-0.705
2dT	C5M	36	0.302	0.436	0.753	10dT	C5M	197	0.129	0.505	-0.729
2dT	C2*	37	0.782	0.378	0.508	10dT	C2*	198	-0.225	0.838	-0.495
2dT	C3*	38	0.917	0.435	0.546	10dT	C3*	199	-0.263	0.979	-0.534
2dT	O3*	39	1.025	0.381	0.469	10dT	O3*	200	-0.371	1.032	-0.455
3dA	P	40	1.063	0.433	0.321	11dA	P	201	-0.352	1.086	-0.304
3dA	O1P	41	0.978	0.549	0.285	11dA	O1P	202	-0.208	1.087	-0.269
3dA	O2P	42	1.210	0.448	0.314	11dA	O2P	203	-0.430	1.211	-0.289
3dA	O5*	43	1.017	0.307	0.233	11dA	O5*	204	-0.423	0.968	-0.221
3dA	C5*	44	1.088	0.183	0.243	11dA	C5*	205	-0.565	0.950	-0.229
3dA	C4*	45	1.010	0.072	0.176	11dA	C4*	206	-0.606	0.819	-0.164
3dA	O4*	46	0.883	0.059	0.242	11dA	O4*	207	-0.536	0.712	-0.231
3dA	C1*	47	0.783	0.043	0.140	11dA	C1*	208	-0.483	0.625	-0.130
3dA	N9	48	0.655	0.095	0.194	11dA	N9	209	-0.358	0.565	-0.184
3dA	C4	49	0.557	0.019	0.243	11dA	C4	210	-0.350	0.443	-0.237
3dA	N3	50	0.535	-0.113	0.243	11dA	N3	211	-0.435	0.340	-0.244
3dA	C2	51	0.426	-0.161	0.300	11dA	C2	212	-0.398	0.229	-0.305
3dA	N1	52	0.336	-0.083	0.358	11dA	N1	213	-0.279	0.214	-0.360
3dA	C6	53	0.353	0.049	0.362	11dA	C6	214	-0.190	0.313	-0.357
3dA	N6	54	0.263	0.128	0.420	11dA	N6	215	-0.070	0.299	-0.412
3dA	C5	55	0.466	0.104	0.303	11dA	C5	216	-0.224	0.432	-0.294
3dA	N7	56	0.509	0.228	0.288	11dA	N7	217	-0.159	0.546	-0.273
3dA	C8	57	0.623	0.222	0.220	11dA	C8	218	-0.241	0.625	-0.205
3dA	C2*	58	0.834	0.121	0.021	11dA	C2*	219	-0.459	0.712	-0.009
3dA	C3*	59	0.984	0.103	0.029	11dA	C3*	220	-0.570	0.813	-0.017
3dA	O3*	60	1.036	-0.003	-0.052	11dA	O3*	221	-0.688	0.779	0.059
4dC	P	61	1.017	-0.015	-0.212	12dC	P	222	-0.689	0.744	0.217
4dC	O1P	62	0.970	0.113	-0.268	12dC	O1P	223	-0.565	0.792	0.281
4dC	O2P	63	1.140	-0.077	-0.269	12dC	O2P	224	-0.820	0.786	0.273
4dC	O5*	64	0.897	-0.121	-0.218	12dC	O5*	225	-0.684	0.583	0.211
4dC	C5*	65	0.914	-0.252	-0.161	12dC	C5*	226	-0.788	0.511	0.144
4dC	C4*	66	0.783	-0.327	-0.152	12dC	C4*	227	-0.755	0.363	0.130
4dC	O4*	67	0.687	-0.244	-0.082	12dC	O4*	228	-0.627	0.350	0.064
4dC	C1*	68	0.564	-0.253	-0.155	12dC	C1*	229	-0.554	0.249	0.136
4dC	N1	69	0.480	-0.134	-0.124	12dC	N1	230	-0.409	0.266	0.111
4dC	C6	70	0.522	-0.004	-0.157	12dC	C6	231	-0.341	0.382	0.150
4dC	C2	71	0.364	-0.149	-0.047	12dC	C2	232	-0.340	0.172	0.033
4dC	O2	72	0.318	-0.261	-0.024	12dC	O2	233	-0.393	0.064	0.004
4dC	N3	73	0.297	-0.042	-0.004	12dC	N3	234	-0.214	0.195	-0.005
4dC	C4	74	0.337	0.083	-0.032	12dC	C4	235	-0.149	0.306	0.029

4dC	N4	75	0.273	0.185	0.024	12dC	N4	236	-0.028	0.328	-0.023
4dC	C5	76	0.450	0.105	-0.110	12dC	C5	237	-0.210	0.402	0.108
4dC	C2*	77	0.606	-0.262	-0.300	12dC	C2*	238	-0.595	0.269	0.280
4dC	C3*	78	0.725	-0.355	-0.290	12dC	C3*	239	-0.743	0.295	0.265
4dC	O3*	79	0.690	-0.495	-0.297	12dC	O3*	240	-0.823	0.175	0.263
5dG	P	80	0.617	-0.565	-0.423	13dG	P	241	-0.829	0.066	0.382
5dG	O1P	81	0.631	-0.479	-0.543	13dG	O1P	242	-0.779	0.126	0.508
5dG	O2P	82	0.661	-0.706	-0.430	13dG	O2P	243	-0.964	0.005	0.382
5dG	O5*	83	0.463	-0.559	-0.374	13dG	O5*	244	-0.721	-0.042	0.331
5dG	C5*	84	0.420	-0.640	-0.264	13dG	C5*	245	-0.751	-0.124	0.217
5dG	C4*	85	0.269	-0.639	-0.249	13dG	C4*	246	-0.652	-0.239	0.203
5dG	O4*	86	0.224	-0.507	-0.213	13dG	O4*	247	-0.521	-0.187	0.173
5dG	C1*	87	0.113	-0.475	-0.298	13dG	C1*	248	-0.427	-0.250	0.263
5dG	N9	88	0.111	-0.327	-0.314	13dG	N9	249	-0.317	-0.154	0.286
5dG	C4	89	0.023	-0.247	-0.255	13dG	C4	250	-0.196	-0.165	0.233
5dG	N3	90	-0.098	-0.272	-0.201	13dG	N3	251	-0.132	-0.269	0.178
5dG	C2	91	-0.172	-0.172	-0.152	13dG	C2	252	-0.007	-0.257	0.133
5dG	N2	92	-0.293	-0.198	-0.105	13dG	N2	253	0.055	-0.363	0.082
5dG	N1	93	-0.124	-0.041	-0.153	13dG	N1	254	0.059	-0.134	0.140
5dG	C6	94	0.000	-0.014	-0.208	13dG	C6	255	-0.006	-0.025	0.196
5dG	O6	95	0.047	0.100	-0.204	13dG	O6	256	0.047	0.086	0.197
5dG	C5	96	0.074	-0.118	-0.261	13dG	C5	257	-0.135	-0.041	0.244
5dG	N7	97	0.190	-0.124	-0.326	13dG	N7	258	-0.219	0.040	0.307
5dG	C8	98	0.211	-0.251	-0.359	13dG	C8	259	-0.329	-0.030	0.334
5dG	C2*	99	0.138	-0.548	-0.428	13dG	C2*	260	-0.506	-0.282	0.388
5dG	C3*	100	0.200	-0.677	-0.379	13dG	C3*	261	-0.641	-0.320	0.331
5dG	O3*	101	0.105	-0.782	-0.353	13dG	O3*	262	-0.655	-0.460	0.299
6dT	P	102	0.016	-0.851	-0.468	14dT	P	263	-0.654	-0.576	0.411
6dT	O1P	103	0.089	-0.845	-0.597	14dT	O1P	264	-0.707	-0.522	0.539
6dT	O2P	104	-0.032	-0.982	-0.419	14dT	O2P	265	-0.716	-0.697	0.354
6dT	O5*	105	-0.107	-0.748	-0.475	14dT	O5*	266	-0.496	-0.598	0.427
6dT	C5*	106	-0.200	-0.738	-0.366	14dT	C5*	267	-0.419	-0.659	0.322
6dT	C4*	107	-0.298	-0.625	-0.389	14dT	C4*	268	-0.272	-0.657	0.356
6dT	O4*	108	-0.236	-0.495	-0.384	14dT	O4*	269	-0.215	-0.525	0.356
6dT	C1*	109	-0.324	-0.411	-0.461	14dT	C1*	270	-0.100	-0.536	0.442
6dT	N1	110	-0.254	-0.287	-0.504	14dT	N1	271	-0.057	-0.401	0.487
6dT	C6	111	-0.134	-0.293	-0.575	14dT	C6	272	-0.145	-0.316	0.556
6dT	C2	112	-0.299	-0.162	-0.457	14dT	C2	273	0.067	-0.350	0.446
6dT	O2	113	-0.412	-0.147	-0.411	14dT	O2	274	0.155	-0.422	0.399
6dT	N3	114	-0.220	-0.049	-0.478	14dT	N3	275	0.098	-0.216	0.470
6dT	C4	115	-0.099	-0.058	-0.546	14dT	C4	276	0.008	-0.132	0.536
6dT	O4	116	-0.028	0.043	-0.554	14dT	O4	277	0.035	-0.013	0.545
6dT	C5	117	-0.057	-0.180	-0.598	14dT	C5	278	-0.113	-0.184	0.582
6dT	C5M	118	0.066	-0.193	-0.687	14dT	C5M	279	-0.209	-0.102	0.668
6dT	C2*	119	-0.366	-0.495	-0.579	14dT	C2*	280	-0.143	-0.624	0.556
6dT	C3*	120	-0.363	-0.637	-0.526	14dT	C3*	281	-0.247	-0.716	0.494
6dT	O3*	121	-0.494	-0.695	-0.511	14dT	O3*	282	-0.202	-0.852	0.481
7dA	P	122	-0.589	-0.733	-0.636	15dA	P	283	-0.177	-0.949	0.607
7dA	O1P	123	-0.507	-0.741	-0.760	15dA	O1P	284	-0.254	-0.900	0.724
7dA	O2P	124	-0.673	-0.849	-0.599	15dA	O2P	285	-0.196	-1.089	0.563
7dA	O5*	125	-0.680	-0.601	-0.645	15dA	O5*	286	-0.021	-0.922	0.634
7dA	C5*	126	-0.765	-0.567	-0.534	15dA	C5*	287	0.075	-0.956	0.533
7dA	C4*	127	-0.811	-0.423	-0.542	15dA	C4*	288	0.207	-0.886	0.557
7dA	O4*	128	-0.696	-0.337	-0.556	15dA	O4*	289	0.187	-0.744	0.568
7dA	C1*	129	-0.731	-0.236	-0.652	15dA	C1*	290	0.281	-0.699	0.667
7dA	N9	130	-0.607	-0.189	-0.718	15dA	N9	291	0.233	-0.571	0.724
7dA	C4	131	-0.552	-0.070	-0.700	15dA	C4	292	0.291	-0.453	0.706
7dA	N3	132	-0.594	0.041	-0.636	15dA	N3	293	0.405	-0.416	0.646

7dA	C2	133	-0.517	0.148	-0.635	15dA	C2	294	0.439	-0.289	0.646
7dA	N1	134	-0.400	0.152	-0.696	15dA	N1	295	0.366	-0.195	0.702
7dA	C6	135	-0.353	0.045	-0.761	15dA	C6	296	0.251	-0.225	0.763
7dA	N6	136	-0.232	0.048	-0.816	15dA	N6	297	0.176	-0.130	0.819
7dA	C5	137	-0.430	-0.070	-0.764	15dA	C5	298	0.210	-0.358	0.766
7dA	N7	138	-0.414	-0.189	-0.821	15dA	N7	299	0.106	-0.421	0.819
7dA	C8	139	-0.523	-0.260	-0.794	15dA	C8	300	0.121	-0.550	0.794
7dA	C2*	140	-0.825	-0.304	-0.749	15dA	C2*	301	0.287	-0.809	0.769
7dA	C3*	141	-0.901	-0.401	-0.662	15dA	C3*	302	0.270	-0.934	0.687
7dA	O3*	142	-1.028	-0.350	-0.617	15dA	O3*	303	0.394	-1.001	0.656
8dC	P	143	-1.156	-0.336	-0.715	16dC	P	304	0.477	-1.086	0.765
8dC	O1P	144	-1.140	-0.426	-0.831	16dC	O1P	305	0.389	-1.121	0.879
8dC	O2P	145	-1.278	-0.348	-0.632	16dC	O2P	306	0.549	-1.194	0.694
8dC	O5*	146	-1.139	-0.185	-0.766	16dC	O5*	307	0.582	-0.975	0.816
8dC	C5*	147	-1.167	-0.074	-0.680	16dC	C5*	308	0.692	-0.934	0.734
8dC	C4*	148	-1.108	0.054	-0.738	16dC	C4*	309	0.752	-0.807	0.790
8dC	O4*	149	-0.966	0.043	-0.752	16dC	O4*	310	0.655	-0.701	0.791
8dC	C1*	150	-0.930	0.146	-0.845	16dC	C1*	311	0.707	-0.605	0.884
8dC	N1	151	-0.809	0.106	-0.921	16dC	N1	312	0.593	-0.533	0.948
8dC	C6	152	-0.801	-0.017	-0.987	16dC	C6	313	0.487	-0.602	1.009
8dC	C2	153	-0.699	0.192	-0.927	16dC	C2	314	0.589	-0.393	0.947
8dC	O2	154	-0.701	0.302	-0.872	16dC	O2	315	0.679	-0.328	0.895
8dC	N3	155	-0.589	0.156	-0.994	16dC	N3	316	0.487	-0.328	1.004
8dC	C4	156	-0.579	0.038	-1.056	16dC	C4	317	0.385	-0.392	1.062
8dC	N4	157	-0.465	0.006	-1.117	16dC	N4	318	0.288	-0.322	1.118
8dC	C5	158	-0.685	-0.051	-1.055	16dC	C5	319	0.383	-0.531	1.066
8dC	C2*	159	-1.050	0.169	-0.936	16dC	C2*	320	0.792	-0.679	0.986
8dC	C3*	160	-1.160	0.084	-0.876	16dC	C3*	321	0.795	-0.822	0.934
8dC	O3*	161	-1.277	0.167	-0.872	16dC	O3*	322	0.929	-0.872	0.946

Appendix 4 Experimental restraints and coordinates for CATG[a]

A. NOE intensities and distances

Mixing time (sec.)				Distance estimate (nm.)		Proton pair
0.100	0.200	0.500	0.500	Lower	Upper	
1. Resolved cross-peaks						
0.130	0.267	0.600	0.294	0.300	0.330	1dC 1' - 1dC 2'
0.800	1.000	1.500	1.002	0.220	0.250	1dC 1' - 1dC 2"
0.000	0.190	0.450	0.989			1dC 1' - 1dC 4'
0.000	0.054	0.145	0.258			1dC 1' - 1dC 3'
0.056	0.176	0.478	0.588	0.390	0.440	1dC 1' - 1dC H6
0.000	0.064	0.077	0.111	0.400	0.430	1dC 1' - 2dA H8
0.000	0.247	0.295	0.298	0.290	0.380	1dC 2" - 1dC H6
0.282	0.327	0.363	0.446	0.260	0.300	1dC 2" - 2dA H8
0.735	1.038	1.334	1.283	0.220	0.260	1dC 2' - 1dC H6
0.000	0.000	0.053	0.161			1dC 2' - 1dC H5
0.309	0.905	1.273	0.772	0.250	0.300	1dC 2' - 1dC 3'
0.084	0.302	0.323	0.363	0.320	0.390	1dC 2' - 2dA H8
0.129	0.224	0.471	0.565	0.290	0.340	1dC 3' - 1dC 2"
		0.271	0.620			1dC 3' - 1dC 4'
0.122	0.188	0.390	0.349	0.300	0.360	1dC 3' - 1dC H6
0.000			0.225[b]			1dC 4' - 1dC H6
0.000	0.000	0.014	0.225			1dC 3' - 2dA H8
0.000	0.000	0.005	0.105			1dC H6 - 2dA H8
0.041	0.085	0.424	0.529	0.360	0.430	2dA 1' - 2dA 3'
0.309	0.891	1.101	1.016	0.270	0.300	2dA 1' - 2dA 2'
0.000	0.178	0.501	0.368	0.340	0.400	2dA 1' - 2dA H8
1.086	1.763	1.765	1.439	0.210	0.230	2dA 1' - 2dA 2"
0.000	0.067	0.260	0.579	0.430	0.530	2dA 1' - 3dT H6
0.000	0.083	0.183	0.326	0.490	0.590	2dA 1' - 3dT HM
	0.212[b]		0.428[b]			2dA 2' - 3dT HM
0.000	0.000	0.000	0.428[b]			2dA 3' - 3dT HM
0.279	0.905	1.076	0.790	0.260	0.310	2dA 2" - 2dA 3'
0.000	0.140	0.229	0.174	0.320	0.400	2dA 2" - 2dA 4'
0.288	0.423	0.968	0.989	0.270	0.360	2dA 2" - 2dA H8
0.606	0.750	0.683	0.772	0.230	0.250	2dA 2" - 3dT H6
0.000	0.158[b]	0.346	0.542	0.350	0.450	2dA 2" - 3dT HM
0.000	0.082	0.397	0.193	0.340	0.500	2dA 2' - 2dA 4'
1.315	1.872	1.857	1.407	0.200	0.220	2dA 2' - 2dA H8
0.672	1.727	1.570	1.369	0.225	0.260	2dA 2' - 2dA 3'
	0.343	0.342	0.492	0.280	0.360	2dA 2' - 3dT H6
0.043	0.074	0.424	0.469	0.360	0.410	2dA 3' - 2dA H8
0.000		0.085	0.239			2dA 4' - 2dA H8
0.000	0.129	0.115	0.354	0.330	0.450	2dA 3' - 3dT H6
0.672	1.040	1.500	0.310	0.230	0.260	2dA 4' - 2dA 3'
0.191	0.411	0.647	0.819	0.300	0.330	2dA 4' - 2dA 1'
0.598	1.747	2.241	1.947	0.230	0.270	2dA H8 - 3dT HM
0.000		0.132	0.212			2dA H8 - 3dT H6
0.000		0.183	0.082			2dA H2 -16dG 1'

0.800	1.430	1.700	1.466	0.215	0.250	3dT 1' - 3dT 2"
0.000	0.129	0.323	0.487	0.370	0.430	3dT 1' - 4dG H8
0.140	0.627	0.838	1.012	0.260	0.380	3dT 2" - 3dT H6
0.296	0.466	0.436	0.625	0.260	0.290	3dT 2" - 4dG H8
0.130	0.483	0.714	0.676	0.290	0.340	3dT 2' - 3dT 1'
1.313	1.712	1.885	1.411	0.200	0.220	3dT 2' - 3dT H6
0.292	1.095		0.593			3dT 4' - 3dT 3'
0.617	0.970	1.059	0.869	0.230	0.260	3dT 2' - 3dT 3'
0.083	0.179	0.187	0.473	0.330	0.380	3dT 2' - 4dG H8
0.000		0.045	0.092			3dT 3' - 4dG H8
0.000		0.075	0.189			3dT H6 - 4dG H8
0.146	0.376	0.513	0.997	0.300	0.340	4dG 1' - 4dG 4'
0.000	0.130	0.285	0.672	0.390	0.450	4dG 1' - 4dG 3'
0.361	0.631	1.108	0.607	0.250	0.300	4dG 1' - 4dG 2'
0.035	0.101	0.321	0.405	0.400	0.450	4dG 1' - 4dG H8
0.881	1.467	1.770	1.361	0.215	0.260	4dG 1' - 4dG 2"
0.123	0.239	0.400	0.556	0.340	0.360	4dG 1' - 5dC H6
0.000		0.163	0.294			4dG 1' - 5dC H5
0.422	0.748	0.773	0.887	0.250	0.270	4dG 2" - 5dC H6
0.111	0.312	0.485	0.409	0.300	0.360	4dG 2" - 5dC H5
0.000			0.336[b]			4dG 2" - 4dG 4'
0.000			0.271[b]			4dG 2' - 4dG 4'
0.145	0.221	0.511	0.542	0.300	0.370	4dG 2' - 5dC H6
0.000	0.174	0.268	0.313	0.330	0.400	4dG 2' - 5dC H5
0.000	0.079	0.365	0.464	0.430	0.500	4dG 3' - 4dG H8
0.000			0.244			4dG 4' - 4dG H8
0.000		0.120	0.281			4dG 3' - 5dC H6
0.000	0.000	0.038	0.198			4dG H8 - 5dC H6
0.779	0.900	2.006	1.273	0.190	0.210	4dG H8 - 4dG 2'
0.199	0.583	0.916	0.979	0.280	0.350	4dG H8 - 4dG 2"
0.000	0.090	0.170	0.258	0.370	0.450	4dG H8 - 5dC H5
0.196	0.517	0.769	0.469	0.280	0.320	5dC 1' - 5dC 2'
0.000	0.064	0.381	0.515	0.360	0.450	5dC 1' - 5dC H6
0.154	0.629[b]	1.136	2.251[b]	0.290	0.360	5dC 1' - 5dC 4'
0.000		0.319	0.151			5dC 1' - 5dC 3'
0.292	1.095		1.838			5dC 4' - 5dC 3'
0.905	1.392	1.521	1.144	0.215	0.240	5dC 1' - 5dC 2"
0.073	0.173	0.278	0.340	0.360	0.390	5dC 1' - 6dA H8
0.169	0.526	0.906	1.034	0.290	0.350	5dC 2" - 5dC H6
0.236	0.464	0.554	0.473	0.270	0.310	5dC 2" - 6dA H8
0.000	0.000	0.231	0.248	0.390	0.450	5dC 2' - 5dC H5
	0.067	0.121	0.166			5dC 2" - 5dC H5
0.617	0.970	1.059	0.699	0.230	0.260	5dC 2' - 5dC 3'
0.177	0.173	0.344	0.345	0.290	0.360	5dC 2' - 6dA H8
0.000	0.000	0.000	0.088[b]			5dC 3' - 6dA H8
0.217	0.268	0.289	0.625	0.270	0.360	5dC 3' - 5dC H6
1.490	1.920	1.582	1.617	0.190	0.210	5dC H6 - 5dC 2'
0.000		0.097	0.189			5dC H6 - 6dA H8
0.935	1.683	1.862	1.540	0.215	0.240	6dA 1' - 6dA 2"
0.275	0.752	1.092	1.168	0.280	0.310	6dA 1' - 6dA 2'
0.000	0.058		0.676			6dA 1' - 6dA 3'
0.045	0.252		0.616	0.340	0.400	6dA 1' - 6dA 4'
0.000	0.000		0.111			6dA 1' - 6dA H2
0.000	0.139	0.333	0.450	0.350	0.420	6dA 1' - 6dA H8
0.000	0.000		0.202			6dA 1' - 7dT 1'
0.000	0.117	0.222	0.473	0.380	0.430	6dA 1' - 7dT H6
0.000		0.199	0.349			6dA 1' - 7dT HM
0.203	0.428	0.998	1.034	0.300	0.400	6dA 2" - 6dA H8

0.504	0.750	0.610	0.666	0.240	0.260	6dA 2" - 7dT H6
0.955	1.712	1.678	1.287	0.210	0.230	6dA 2' - 6dA H8
0.199	0.270	0.492	0.515	0.280	0.330	6dA 2' - 7dT H6
0.000	0.121	0.262	0.662			6dA 2' - 7dT HM
		0.309	0.547			6dA 2" - 7dT HM
0.350	0.876	1.102	0.997	0.250	0.290	6dA 3' - 6dA 2"
0.048	0.061	0.364	0.542	0.360	0.460	6dA 3' - 6dA H8
0.000			0.460			6dA 3' - 7dT H6
0.000			0.428[b]			6dA 3' - 7dT HM
0.000	0.000	0.249	0.143			6dA 4' - 6dA 2"
0.000			0.239[b]			6dA 4' - 6dA 2'
0.000	0.000	0.070	0.244			6dA 4' - 6dA H8
0.000	0.000	0.072	0.115			6dA H2 - 7dT 1'
0.000	0.000		0.202			6dA H8 - 7dT H6
0.411	0.563	0.810	0.896	0.310	0.350	6dA H8 - 7dT HM
1.075	1.794	1.870	1.516	0.210	0.235	7dT 1' - 7dT 2"
0.192	0.534	1.009	0.722	0.290	0.330	7dT 1' - 7dT 2'
0.069	0.086	0.088	0.202	0.350	0.420	7dT 1' - 8dG H8
0.226	0.833	0.584	0.827	0.240	0.290	7dT 2" - 8dG H8
0.503	0.447[b]	1.088	0.253[b]	0.230	0.270	7dT 2' - 7dT 3'
0.215	0.312	0.285	0.354	0.270	0.320	7dT 2' - 8dG H8
0.272	0.357[b]	0.650	0.281[b]	0.260	0.290	7dT 3' - 7dT 2"
0.000	0.000	0.000	0.046[b]			7dT 3' - 8dG H8
0.000			0.235[b]			7dT 3' - 7dT HM
0.121	0.447	0.861	1.016	0.330	0.390	7dT H6 - 7dT 2"
1.072	1.680	1.715	1.466	0.210	0.230	7dT H6 - 7dT 2'
0.000		0.075	0.138			7dT H6 - 8dG H8
0.880	1.553	2.171	2.123	0.215	0.260	8dG 1' - 8dG 2"
0.272	0.569	0.911	1.241	0.260	0.300	8dG 1' - 8dG 2'
0.000	0.287	0.366	0.510	0.290	0.360	8dG 1' - 8dG 3'
0.000	0.155	0.520	1.016			8dG 1' - 8dG 4'
0.012	0.086	0.294	0.313	0.400	0.500	8dG 1' - 8dG H8
0.357	0.555	1.283	1.250	0.240	0.280	8dG 2' - 8dG 3'
0.779	0.900	2.006	1.857	0.210	0.270	8dG 2' - 8dG H8
0.025		0.584	1.356[b]			8dG 2" - 8dG H8
	0.947	1.001	1.006	0.220	0.260	8dG 3' - 8dG 2"
			0.937			8dG 3' - 8dG 4'
0.000	0.093	0.386	0.418	0.370	0.500	8dG 3' - 8dG H8

2. Overlapping cross-peaks

0.712	1.700	2.230	2.205	0.220	0.330	1dC 4' - 1dC 5's
	0.165	0.350	0.434	0.380	0.500	1dC H6 - 1dC 5's
0.242	0.868[b]	1.300	1.510[b]			3dT 1' - 3dT 4'
						3dT 1' - 4dG 5'[c]
0.093	0.238	0.355	0.294	0.320	0.360	3dT 1' - 3dT H6
						7dT 1' - 7dT H6[d]
0.376	0.680	0.990	0.556	0.250	0.270	3dT 3' - 3dT 2"
0.376	0.680[b]	0.990	0.652[b]	0.250	0.270	5dC 3' - 5dC 2"
0.127	0.188	0.416	0.616	0.300	0.370	3dT 3' - 3dT H6
0.127	0.188	0.114	0.786	0.300	0.370	7dT 3' - 7dT H6
	0.090[b]	0.259	0.230[b]			3dT 4' - 3dT H6
						7dT 4' - 7dT H6[d]
0.318	0.700[b]	1.500	0.684[b]	0.250	0.290	4dG 3' - 4dG 4'
						6dA 3' - 6dA 4'[d]

3. Covalently bound proton pairs

2.801	3.632[b]	3.082	2.201[b]	0.176	1dC 2' - 1dC 2"
					3dT 2' - 3dT 2"[d]
					5dC 2' - 5dC 2"[d]
0.761	1.412	2.105	2.749	0.246	1dC H6 - 1dC H5
3.762	4.429	3.054	2.201	0.176	2dA 2' - 2dA 2"
0.803	1.445	2.039	1.793	0.295	3dT H6 - 3dT HM
					7dT H6 - 7dT HM[d]
			2.680[b]	0.176	4dG 2' - 4dG 2"
1.027	1.619	2.047	2.547	0.246	5dC H6 - 5dC H5
3.531	4.913	3.106	2.427	0.176	6dA 2' - 6dA 2"
2.816	4.007	3.001	2.386	0.176	7dT 2' - 7dT 2"
2.900	5.025	4.341	3.300	0.176	8dG 2' - 8dG 2"

[a]The absence of an entry indicates that no distance estimate was made, or that the NOE intensity was not tabulated. 5' and 5" NOE intensities and distances were not used stereospecifically because the ratio of $J_{5'5"}$ to the difference in chemical shifts is often not small, leading to second order effects. 5's refers to non-stereospecifc NOEs to the protons on 5' methylene carbon. NOE intensities are given for one of the two symmetry related strands. [b]NOE noted in subsequent analysis of the data, but not used in structure refinement. [c]NOE and distances are included with the previous entry. [d]NOE and distances are approximately equal to the previous entry.

B. Glycosidic dihedral angle restraints

	Allowed ranges of dihedral angles									
	ν_0		ν_1		ν_2		ν_3		ν_4	
Residue										
1	-40	-19	10	40	-39	01	-31	28	-06	39
2	-40	-19	29	40	-39	-07	-18	28	-06	35
3	-32	-19	33	40	-39	-14	-11	28	-06	32
4	-32	-19	33	40	-39	-14	-11	28	-06	32
5	-40	-19	04	40	-39	20	-36	28	-06	40
6	-40	-19	10	40	-39	02	-31	28	-06	39
7	-40	-19	10	40	-39	01	-31	28	-06	39
8	-40	-19	10	40	-39	01	-31	28	-06	39

C. Coordinates (in nm.) for $[d(C\text{-}A\text{-}T\text{-}G\text{-}C\text{-}A\text{-}T\text{-}G)]_2$
322 atoms, protons omitted

Atom	No.	x	y	z	Atom	No.	x	y	z
1dC O5*	1	0.435	0.697	1.275	9dC O5*	162	0.246	0.796	-1.275
1dC C5*	2	0.576	0.686	1.299	9dC C5*	163	0.138	0.886	-1.302
1dC C4*	3	0.639	0.586	1.203	9dC C4*	164	0.023	0.862	-1.204
1dC O4*	4	0.572	0.459	1.214	9dC O4*	165	-0.023	0.726	-1.213
1dC C1*	5	0.584	0.397	1.085	9dC C1*	166	-0.082	0.698	-1.085
1dC N1	6	0.471	0.302	1.070	9dC N1	167	-0.081	0.550	-1.065
1dC C6	7	0.337	0.343	1.083	9dC C6	168	0.038	0.476	-1.076
1dC C2	8	0.498	0.165	1.061	9dC C2	169	-0.202	0.481	-1.055
1dC O2	9	0.612	0.125	1.037	9dC O2	170	-0.306	0.541	-1.029
1dC N3	10	0.397	0.077	1.067	9dC N3	171	-0.203	0.348	-1.061
1dC C4	11	0.270	0.114	1.080	9dC C4	172	-0.091	0.275	-1.073

1dC	N4	12	0.174	0.022	1.085	9dC	N4	173	-0.098	0.143	-1.083
1dC	C5	13	0.236	0.249	1.087	9dC	C5	174	0.033	0.338	-1.080
1dC	C2*	14	0.577	0.510	0.983	9dC	C2*	175	0.004	0.770	-0.983
1dC	C3*	15	0.624	0.633	1.059	9dC	C3*	176	0.066	0.885	-1.060
1dC	O3*	16	0.752	0.680	1.012	9dC	O3*	177	0.016	1.012	-1.017
2dA	P	17	0.777	0.734	0.862	10dA	P	178	0.038	1.071	-0.868
2dA	O1P	18	0.649	0.777	0.802	10dA	O1P	179	0.156	1.007	-0.806
2dA	O2P	19	0.890	0.829	0.865	10dA	O2P	180	0.033	1.219	-0.875
2dA	O5*	20	0.823	0.598	0.789	10dA	O5*	181	-0.094	1.016	-0.793
2dA	C5*	21	0.946	0.534	0.829	10dA	C5*	182	-0.223	1.064	-0.832
2dA	C4*	22	0.966	0.403	0.755	10dA	C4*	183	-0.333	0.993	-0.755
2dA	O4*	23	0.855	0.315	0.783	10dA	O4*	184	-0.327	0.852	-0.782
2dA	C1*	24	0.821	0.250	0.659	10dA	C1*	185	-0.351	0.784	-0.658
2dA	N9	25	0.677	0.215	0.666	10dA	N9	186	-0.282	0.652	-0.666
2dA	C4	26	0.629	0.091	0.674	10dA	C4	187	-0.343	0.535	-0.675
2dA	N3	27	0.684	-0.029	0.656	10dA	N3	188	-0.470	0.496	-0.660
2dA	C2	28	0.609	-0.138	0.670	10dA	C2	189	-0.502	0.369	-0.675
2dA	N1	29	0.481	-0.132	0.702	10dA	N1	190	-0.413	0.277	-0.705
2dA	C6	30	0.421	-0.015	0.721	10dA	C6	191	-0.284	0.309	-0.722
2dA	N6	31	0.294	-0.009	0.761	10dA	N6	192	-0.194	0.217	-0.754
2dA	C5	32	0.495	0.102	0.706	10dA	C5	193	-0.246	0.441	-0.706
2dA	N7	33	0.464	0.230	0.714	10dA	N7	194	-0.129	0.503	-0.713
2dA	C8	34	0.575	0.299	0.688	10dA	C8	195	-0.152	0.632	-0.687
2dA	C2*	35	0.851	0.351	0.551	10dA	C2*	196	-0.293	0.874	-0.551
2dA	C3*	36	0.971	0.426	0.605	10dA	C3*	197	-0.317	1.013	-0.605
2dA	O3*	37	1.096	0.376	0.553	10dA	O3*	198	-0.435	1.075	-0.551
3dT	P	38	1.148	0.416	0.406	11dT	P	199	-0.437	1.141	-0.404
3dT	O1P	39	1.089	0.545	0.366	11dT	O1P	200	-0.299	1 179	-0.365
3dT	O2P	40	1.296	0.401	0.404	11dT	O2P	201	-0.543	1.244	-0.401
3dT	O5*	41	1.081	0.301	0.317	11dT	O5*	202	-0.481	1.015	-0.314
3dT	C5*	42	1.131	0.166	0.322	11dT	C5*	203	-0.615	0.964	-0.322
3dT	C4*	43	1.032	0.071	0.257	11dT	C4*	204	-0.625	0.829	-0.253
3dT	O4*	44	0.904	0.076	0.323	11dT	O4*	205	-0.538	0.733	-0.317
3dT	C1*	45	0.809	0.042	0.221	11dT	C1*	206	-0.500	0.641	-0.213
3dT	N1	46	0.674	0.088	0.263	11dT	N1	207	-0.376	0.571	-0.256
3dT	C6	47	0.647	0.225	0.285	11dT	C6	208	-0.257	0.641	-0.279
3dT	C2	48	0.574	-0.005	0.293	11dT	C2	209	-0.380	0.434	-0.286
3dT	O2	49	0.588	-0.125	0.269	11dT	O2	210	-0.479	0.365	-0.263
3dT	N3	50	0.452	0.039	0.345	11dT	N3	211	-0.266	0.372	-0.338
3dT	C4	51	0.427	0.174	0.367	11dT	C4	212	-0.149	0.443	-0.361
3dT	O4	52	0.318	0.208	0.413	11dT	O4	213	-0.051	0.384	-0.407
3dT	C5	53	0.526	0.268	0.336	11dT	C5	214	-0.145	0.579	-0.332
3dT	C5M	54	0.509	0.417	0.367	11dT	C5M	215	-0.024	0.667	-0.365
3dT	C2*	55	0.859	0.111	0.096	11dT	C2*	216	-0.480	0.728	-0.090
3dT	C3*	56	1.010	0.106	0.111	11dT	C3*	217	-0.584	0.838	-0.107
3dT	O3*	57	1.072	0.003	0.031	11dT	O3*	218	-0.701	0.818	-0.026
4dG	P	58	1.060	-0.008	-0.130	12dG	P	219	-0.700	0.799	0.135
4dG	O1P	59	1.025	0.124	-0.187	12dG	O1P	220	-0.577	0.859	0.192
4dG	O2P	60	1.179	-0.080	-0.181	12dG	O2P	221	-0.831	0.841	0.187
4dG	O5*	61	0.931	-0.104	-0.142	12dG	O5*	222	-0.687	0.639	0.145
4dG	C5*	62	0.937	-0.237	-0.088	12dG	C5*	223	-0.792	0.556	0.091
4dG	C4*	63	0.803	-0.308	-0.096	12dG	C4*	224	-0.757	0.409	0.099
4dG	O4*	64	0.703	-0.231	-0.026	12dG	O4*	225	-0.633	0.385	0.029
4dG	C1*	65	0.585	-0.231	-0.108	12dG	C1*	226	-0.556	0.294	0.111
4dG	N9	66	0.504	-0.112	-0.075	12dG	N9	227	-0.412	0.310	0.078
4dG	C4	67	0.385	-0.117	-0.015	12dG	C4	228	-0.339	0.217	0.017
4dG	N3	68	0.303	-0.219	0.011	12dG	N3	229	-0.363	0.088	-0.009
4dG	C2	69	0.182	-0.198	0.065	12dG	C2	230	-0.269	0.010	-0.064

4dG	N2	70	0.100	-0.300	0.082	12dG	N2	231	-0.291	-0.120	-0.079
4dG	N1	71	0.142	-0.069	0.097	12dG	N1	232	-0.144	0.064	-0.097
4dG	C6	72	0.226	0.039	0.073	12dG	C6	233	-0.118	0.198	-0.072
4dG	O6	73	0.195	0.153	0.106	12dG	O6	234	-0.011	0.248	-0.107
4dG	C5	74	0.349	0.014	0.014	12dG	C5	235	-0.217	0.275	-0.013
4dG	N7	75	0.446	0.094	-0.028	12dG	N7	236	-0.219	0.400	0.031
4dG	C8	76	0.539	0.017	-0.084	12dG	C8	237	-0.338	0.420	0.087
4dG	C2*	77	0.638	-0.230	-0.250	12dG	C2*	238	-0.590	0.334	0.253
4dG	C3*	78	0.756	-0.324	-0.240	12dG	C3*	239	-0.738	0.362	0.242
4dG	O3*	79	0.823	-0.462	-0.261	12dG	O3*	240	-0.822	0.248	0.264
5dC	P	80	0.655	-0.521	-0.395	13dC	P	241	-0.821	0.156	0.397
5dC	O1P	81	0.672	-0.426	-0.507	13dC	O1P	242	-0.759	0.229	0.509
5dC	O2P	82	0.700	-0.661	-0.412	13dC	O2P	243	-0.956	0.099	0.415
5dC	O5*	83	0.500	-0.521	-0.352	13dC	O5*	244	-0.720	0.039	0.351
5dC	C5*	84	0.454	-0.606	-0.245	13dC	C5*	245	-0.757	-0.052	0.246
5dC	C4*	85	0.304	-0.596	-0.225	13dC	C4*	246	-0.652	-0.159	0.227
5dC	O4*	86	0.268	-0.461	-0.192	13dC	O4*	247	-0.525	-0.099	0.193
5dC	C1*	87	0.152	-0.426	-0.269	13dC	C1*	248	-0.424	-0.164	0.272
5dC	N1	88	0.154	-0.279	-0.292	13dC	N1	249	-0.313	-0.067	0.294
5dC	C6	89	0.262	-0.218	-0.358	13dC	C6	250	-0.336	0.055	0.359
5dC	C2	90	0.056	-0.197	-0.236	13dC	C2	251	-0.187	-0.090	0.237
5dC	O2	91	-0.050	-0.246	-0.194	13dC	O2	252	-0.156	-0.202	0.196
5dC	N3	92	0.069	-0.065	-0.241	13dC	N3	253	-0.094	0.006	0.242
5dC	C4	93	0.173	-0.004	-0.300	13dC	C4	254	-0.115	0.124	0.300
5dC	N4	94	0.186	0.127	-0.291	13dC	N4	255	-0.023	0.219	0.291
5dC	C5	95	0.271	-0.080	-0.362	13dC	C5	256	-0.236	0.151	0.362
5dC	C2*	96	0.167	-0.503	-0.399	13dC	C2*	257	-0.493	-0.202	0.401
5dC	C3*	97	0.230	-0.633	-0.352	13dC	C3*	258	-0.631	-0.240	0.354
5dC	O3*	98	0.134	-0.736	-0.322	13dC	O3*	259	-0.645	-0.379	0.323
6dA	P	99	0.022	-0.790	-0.426	14dA	P	260	-0.615	-0.500	0.426
6dA	O1P	100	0.067	-0.765	-0.565	14dA	O1P	261	-0.625	-0.450	0.565
6dA	O2P	101	-0.015	-0.927	-0.386	14dA	O2P	262	-0.698	-0.616	0.387
6dA	O5*	102	-0.100	-0.690	-0.392	14dA	O5*	263	-0.460	-0.531	0.394
6dA	C5*	103	-0.162	-0.694	-0.263	14dA	C5*	264	-0.424	-0.582	0.265
6dA	C4*	104	-0.285	-0.606	-0.254	14dA	C4*	265	-0.276	-0.617	0.256
6dA	O4*	105	-0.250	-0.468	-0.277	14dA	O4*	266	-0.195	-0.500	0.279
6dA	C1*	106	-0.345	-0.413	-0.371	14dA	C1*	267	-0.091	-0.536	0.372
6dA	N9	107	-0.280	-0.302	-0.444	14dA	N9	268	-0.050	-0.413	0.445
6dA	C4	108	-0.310	-0.174	-0.427	14dA	C4	269	0.067	-0.352	0.428
6dA	N3	109	-0.417	-0.110	-0.378	14dA	N3	270	0.186	-0.391	0.379
6dA	C2	110	-0.419	0.022	-0.375	14dA	C2	271	0.286	-0.306	0.376
6dA	N1	111	-0.317	0.096	-0.416	14dA	N1	272	0.276	-0.181	0.419
6dA	C6	112	-0.207	0.038	-0.465	14dA	C6	273	0.160	-0.136	0.468
6dA	N6	113	-0.099	0.110	-0.491	14dA	N6	274	0.145	-0.008	0.502
6dA	C5	114	-0.202	-0.100	-0.473	14dA	C5	275	0.052	-0.223	0.475
6dA	N7	115	-0.113	-0.185	-0.522	14dA	N7	276	-0.071	-0.211	0.524
6dA	C8	116	-0.162	-0.307	-0.506	14dA	C8	277	-0.131	-0.328	0.507
6dA	C2*	117	-0.382	-0.530	-0.460	14dA	C2*	278	-0.155	-0.640	0.462
6dA	C3*	118	-0.388	-0.644	-0.359	14dA	C3*	279	-0.236	-0.719	0.362
6dA	O3*	119	-0.517	-0.656	-0.297	14dA	O3*	280	-0.160	-0.824	0.299
7dT	P	120	-0.643	-0.721	-0.372	15dT	P	281	-0.124	-0.962	0.373
7dT	O1P	121	-0.596	-0.815	-0.476	15dT	O1P	282	-0.223	-0.989	0.479
7dT	O2P	122	-0.738	-0.771	-0.269	15dT	O2P	283	-0.100	-1.066	0.269
7dT	O5*	123	-0.705	-0.592	-0.445	15dT	O5*	284	0.016	-0.924	0.444
7dT	C5*	124	-0.784	-0.497	-0.372	15dT	C5*	285	0.137	-0.917	0.367
7dT	C4*	125	-0.806	-0.372	-0.455	15dT	C4*	286	0.247	-0.852	0.449
7dT	O4*	126	-0.681	-0.308	-0.490	15dT	O4*	287	0.215	-0.717	0.487
7dT	C1*	127	-0.715	-0.229	-0.605	15dT	C1*	288	0.298	-0.693	0.601

7dT	N1	128	-0.591	-0.186	-0.674	15dT	N1	289	0.251	-0.571	0.673
7dT	C6	129	-0.499	-0.279	-0.725	15dT	C6	290	0.120	-0.564	0.724
7dT	C2	130	-0.560	-0.050	-0.680	15dT	C2	291	0.333	-0.458	0.679
7dT	O2	131	-0.640	0.037	-0.648	15dT	O2	292	0.451	-0.461	0.645
7dT	N3	132	-0.437	-0.010	-0.732	15dT	N3	293	0.282	-0.340	0.733
7dT	C4	133	-0.346	-0.103	-0.782	15dT	C4	294	0.152	-0.333	0.783
7dT	O4	134	-0.236	-0.064	-0.820	15dT	O4	295	0.109	-0.225	0.823
7dT	C5	135	-0.378	-0.238	-0.780	15dT	C5	296	0.072	-0.446	0.780
7dT	C5M	136	-0.290	-0.344	-0.846	15dT	C5M	297	-0.067	-0.450	0.844
7dT	C2*	137	-0.803	-0.320	-0.688	15dT	C2*	298	0.288	-0.820	0.683
7dT	C3*	138	-0.877	-0.404	-0.585	15dT	C3*	299	0.274	-0.928	0.578
7dT	O3*	139	-1.015	-0.368	-0.572	15dT	O3*	300	0.395	-1.005	0.562
8dG	P	140	-1.124	-0.381	-0.691	16dG	P	301	0.460	-1.096	0.679
8dG	O1P	141	-1.083	-0.488	-0.784	16dG	O1P	302	0.355	-1.140	0.773
8dG	O2P	142	-1.259	-0.387	-0.629	16dG	O2P	303	0.546	-1.198	0.615
8dG	O5*	143	-1.105	-0.239	-0.765	16dG	O5*	304	0.553	-0.988	0.754
8dG	C5*	144	-1.139	-0.117	-0.697	16dG	C5*	305	0.667	-0.932	0.688
8dG	C4*	145	-1.092	0.003	-0.776	16dG	C4*	306	0.725	-0.815	0.766
8dG	O4*	146	-0.949	-0.005	-0.793	16dG	O4*	307	0.622	-0.715	0.783
8dG	C1*	147	-0.924	0.043	-0.926	16dG	C1*	308	0.642	-0.664	0.916
8dG	N9	148	-0.789	0.002	-0.972	16dG	N9	309	0.523	-0.590	0.963
8dG	C4	149	-0.691	0.088	-0.993	16dG	C4	310	0.524	-0.460	0.987
8dG	N3	150	-0.688	0.222	-0.989	16dG	N3	311	0.622	-0.369	0.983
8dG	C2	151	-0.576	0.289	-1.016	16dG	C2	312	0.600	-0.242	1.014
8dG	N2	152	-0.576	0.422	-1.012	16dG	N2	313	0.700	-0.155	1.009
8dG	N1	153	-0.459	0.219	-1.049	16dG	N1	314	0.472	-0.201	1.052
8dG	C6	154	-0.460	0.080	-1.054	16dG	C6	315	0.368	-0.293	1.057
8dG	O6	155	-0.360	0.017	-1.084	16dG	O6	316	0.253	-0.259	1.085
8dG	C5	156	-0.579	0.014	-1.025	16dG	C5	317	0.395	-0.425	1.024
8dG	N7	157	-0.612	-0.114	-1.023	16dG	N7	318	0.320	-0.535	1.021
8dG	C8	158	-0.741	-0.121	-0.991	16dG	C8	319	0.399	-0.635	0.985
8dG	C2*	159	-1.037	-0.014	-1.009	16dG	C2*	320	0.676	-0.785	0.999
8dG	C3*	160	-1.153	0.008	-0.915	16dG	C3*	321	0.771	-0.856	0.905
8dG	O3*	161	-1.212	0.135	-0.940	16dG	O3*	322	0.904	-0.808	0.932

Appendix 5 Experimental restraints and coordinates for L10 DNA[a]

A. NOE intensities and distances in D_2O

Mixing time (sec.)				Distance estimate (nm.)		Proton Pair
0.050	0.100	0.150	0.250	Lower	Upper	
1. Resolved cross-peaks						
0.0090	0.0150	0.0600	0.1090	0.370	0.520	1dT 1' - 2dC H6
0.0000	0.0000	0.0150	0.1280			1dT 1' - 2dC H6
0.0410	0.0980	0.2030	0.3160	0.300	0.380	1dT 2" - 1dT H6
0.3090	0.4540	0.6830	0.7980	0.230	0.240	1dT 2' - 1dT H6
0.0090	0.0280	0.0340	0.1070	0.400	0.470	1dT 3' - 1dT 1'
0.0615	0.0990	0.2170	0.3210	0.280	0.320	1dT 3' - 1dT 2"
0.1850	0.2770	0.4580	0.5860	0.230	0.270	1dT 3' - 1dT 2'
0.0185	0.0315	0.0860	0.1670	0.340	0.420	1dT 3' - 1dT H6
0.0000	0.0300	0.0320	0.0690	0.320	0.440	1dT 3' - 2dC H6
0.1595	0.3330	0.4030	0.7830	0.230	0.280	1dT 4' - 1dT 3'
0.0000	0.0290	0.0520	0.0470	0.320	0.420	1dT 4' - 1dT H6
0.0190	0.0500	0.0925	0.1650	0.380	0.430	1dT H6 - 1dT 1'
0.0080	0.0460	0.0810	0.1280	0.360	0.450	1dT H6 - 2dC H5
0.0000	0.0000	0.0000	0.0160			1dT H6 - 2dC H6
0.0170	0.0550	0.0900	0.1490	0.380	0.430	2dC 1' - 2dC H6
0.0330	0.0270	0.0755	0.1050	0.350	0.400	2dC 1' - 3dT H6
0.0275	0.0200	0.0700	0.1320	0.390	0.500	2dC 1' - 3dT HM
0.0000	0.0000	0.0550	0.0800	0.320	0.460	2dC 1' - 20dA H2
0.4910	0.7280	0.8670[b]	0.9910[b]	0.220	0.260	2dC 2' - 2dC H6
0.0000	0.0000	0.0470	0.1000	0.330	0.500	2dC 3' - 2dC 1'
0.0000	0.0520	0.0710	0.1180	0.290	0.380	2dC 3' - 2dC H6
0.0745	0.1420	0.1560	0.3180	0.250	0.320	2dC 4' - 2dC 1'
0.0000	0.0000	0.0170	0.0550	0.400	0.650	2dC 4' - 2dC H6
0.0000	0.0700	0.1840	0.1880	0.340	0.460	2dC H5 - 3dT HM
0.0000	0.0000	0.0330	0.0330	0.350	0.450	2dC H6 - 3dT H6
0.0720	0.1010	0.1780	0.3660	0.310	0.430	2dC H6 - 3dT HM
0.0000	0.0000	0.0000	0.0260			3dT 1' - 4dA 1'
0.0145	0.0350	0.0705	0.0980	0.390	0.440	3dT 1' - 4dA H8
0.2920	0.5520	0.7060	0.8150	0.210	0.240	3dT 2" - 3dT 1'
0.0690	0.1000	0.1570	0.2360	0.270	0.300	3dT 2" - 4dA H8
0.1025	0.1400	0.2300	0.3970	0.270	0.310	3dT 2' - 3dT 1'
0.6120	0.8790	0.9140[b]	1.1600[b]	0.210	0.240	3dT 2' - 3dT H6
0.0245	0.0300	0.0580	0.1330	0.330	0.390	3dT 2' - 4dA H8
0.0130	0.0410	0.0890	0.1320	0.360	0.430	3dT 3' - 3dT 1'
0.2610	0.3920	0.5900	0.6750	0.210	0.240	3dT 3' - 3dT 2'
0.0120	0.0820	0.0770	0.1750	0.290	0.420	3dT 3' - 3dT H6
0.0000	0.0030	0.0190	0.0440	0.460	0.550	3dT 3' - 4dA H8
0.0000	0.0000	0.0130	0.0450	0.480	0.700	3dT H6 - 4dA H8
0.0135	0.0390	0.0600	0.1390	0.400	0.450	4dA 1' - 4dA H8
0.0000	0.0150	0.0240	0.0320	0.420	0.480	4dA 1' - 5dT 1'
0.0305	0.0600	0.1165	0.1730	0.350	0.380	4dA 1' - 5dT H6
0.0000	0.0000	0.0580	0.1250	0.440	0.550	4dA 1' - 5dT HM
0.2830	0.4110	0.5640	0.6770	0.210	0.240	4dA 2" - 4dA 1'
0.0340	0.1010	0.1310	0.3300	0.270	0.390	4dA 2" - 4dA H8
0.1150	0.2230	0.2030	0.1850	0.230	0.270	4dA 2" - 5dT H6
0.0355	0.1180	0.2090	0.1970	0.330	0.560	4dA 2" - 5dT HM
0.0980	0.1840	0.3320	0.4530	0.260	0.300	4dA 2' - 4dA 1'

0.2770	0.3150	0.4620	0.5180	0.210	0.230	4dA 2' - 4dA H8
0.0465	0.0520	0.1220	0.1310	0.290	0.350	4dA 2' - 5dT H6
0.0105	0.0270	0.0790	0.1850	0.380	0.460	4dA 3' - 4dA 1'
0.1630	0.3080	0.4380	0.5410	0.230	0.270	4dA 3' - 4dA 2"
0.0150	0.0365	0.0695	0.1320	0.360	0.420	4dA 3' - 4dA H8
0.0117	0.0070	0.0340	0.0840	0.360	0.490	4dA 3' - 5dT H6
0.0540	0.1250	0.1780	0.2500	0.280	0.330	4dA 4' - 4dA 1'
0.0150	0.0560	0.0740	0.1170	0.350	0.420	4dA 4' - 4dA 2"
0.0000	0.0000	0.0140	0.0410	0.430	0.650	4dA 4' - 4dA H8
0.0000	0.0000	0.0055[b]	0.0120[b]			4dA H2 - 4dA 1'
0.0000	0.0070	0.0205	0.0480	0.450	0.560	4dA H8 - 5dT H6
0.0735	0.1580	0.2600	0.3930	0.320	0.400	4dA H8 - 5dT HM
0.0230	0.0370	0.0950	0.1420	0.350	0.400	5dT 1' - 5dT H6
0.0000	0.0000	0.0000	0.0290			5dT 1' - 6dC H5
0.0050	0.0370	0.0605	0.1360	0.380	0.550	5dT 1' - 6dC H6
0.3200	0.5360	0.6230	0.9000	0.210	0.240	5dT 2" - 5dT 1'
0.0535	0.1310	0.2700	0.2130	0.280	0.370	5dT 2" - 5dT H6
0.1740	0.2220	0.2170[b]	0.3590[b]	0.260	0.300	5dT 2" - 6dC H6
0.0420	0.1300	0.2160	0.3540	0.300	0.370	5dT 2' - 5dT 1'
0.2625	0.5220	0.6670	0.4330	0.200	0.230	5dT 2' - 5dT H6
0.0270	0.0490	0.0640	0.1540	0.340	0.460	5dT 3' - 5dT 1'
0.0000	0.0510	0.0620	0.1130	0.300	0.420	5dT 3' - 5dT H6
0.0000	0.0000	0.0060[b]	0.0130	0.480	0.670	5dT H6 - 4dA 4'
0.0140	0.0420	0.0630	0.1120	0.370	0.440	5dT H6 - 6dC H5
0.0000	0.0000	0.0100	0.0580		0.480	5dT H6 - 6dC H6
0.0210	0.0350	0.0690	0.1260	0.370	0.410	6dC 1' - 6dC H6
0.0195	0.0320	0.0545	0.0790	0.370	0.410	6dC 1' - 7dA H8
0.2230	0.4550	0.5310[b]	0.6910[b]	0.230	0.300	6dC 2" - 6dC 1'
0.0520	0.1100	0.2830[b]	0.4000[b]	0.330	0.430	6dC 2" - 6dC H6
0.0550	0.1030	0.1240	0.1920	0.280	0.320	6dC 2" - 7dA H8
0.0320	0.0980	0.1950	0.3250	0.290	0.350	6dC 2' - 6dC 1'
0.5360	0.8430	0.8560[b]	1.0240[b]	0.210	0.250	6dC 2' - 6dC H6
0.0485	0.0480	0.1020	0.1050	0.290	0.360	6dC 2' - 7dA H8
0.0370	0.1260	0.1100	0.1910[b]	0.330	0.440	6dC 3' - 6dC H6
0.0000	0.0000	0.0220	0.0310	0.350	0.500	6dC 3' - 7dA H8
0.0000	0.0270	0.0740[b]		0.470	0.700	6dC H5 - 5dT HM
0.0000	0.0000	0.0140	0.0230	0.460	0.530	6dC H6 - 7dA H8
0.0165	0.0470	0.0720	0.1640	0.380	0.440	7dA 1' - 7dA H8
0.0000	0.0000	0.0150	0.0180	0.400	0.540	7dA 1' - 8dC 1'
0.0000	0.0000	0.0150	0.0640			7dA 1' - 8dC H5
0.0215	0.0930	0.1310	0.2140	0.320	0.450	7dA 1' - 8dC H6
0.0775	0.1320	0.1480	0.4320	0.280	0.360	7dA 2" - 7dA H8
0.0400	0.0370	0.1200	0.1710	0.280	0.370	7dA 2" - 8dC H5
0.1600	0.2720	0.3010	0.3070	0.220	0.260	7dA 2" - 8dC H6
0.0835	0.2290	0.3900	0.5950	0.270	0.320	7dA 2' - 7dA 1'
0.3675	0.5070	0.7090	0.7270	0.205	0.230	7dA 2' - 7dA H8
0.0000	0.0600	0.0940	0.1730	0.300	0.360	7dA 2' - 8dC H5
0.0520	0.1050	0.1110	0.2080	0.270	0.350	7dA 2' - 8dC H6
0.0153	0.0470	0.0930	0.1520	0.350	0.420	7dA 3' - 7dA H8
0.0000	0.0000	0.0170	0.0300	0.430	0.540	7dA 3' - 8dC H5
0.0000	0.0130	0.0280	0.0610	0.330	0.500	7dA 3' - 8dC H6
0.0630	0.1360	0.2010	0.2650	0.260	0.360	7dA 4' - 7dA 1'
0.0000	0.0000	0.0440	0.0950	0.360	0.500	7dA 4' - 7dA 2"
0.1845	0.3700	0.6300	0.6700	0.220	0.270	7dA 4' - 7dA 3'
0.0000	0.0000	0.0035	0.0400			7dA 4' - 7dA H8
0.0100	0.0420	0.0590	0.0850	0.370	0.470	7dA H8 - 8dC H5
0.0000	0.0035	0.0275	0.0770	0.480	0.540	7dA H8 - 8dC H6
0.0215	0.0420	0.0880	0.1710	0.360	0.420	8dC 1' - 8dC H6

0.0160	0.0260	0.0435	0.0640	0.390	0.430	8dC 1' - 9dC H6
0.2880	0.5040	0.7180	0.8970	0.210	0.240	8dC 2" - 8dC 1'
0.0000	0.0160	0.0445	0.0840	0.410	0.520	8dC 2" - 8dC H5
0.0435	0.1490	0.2580	0.3560	0.300	0.420	8dC 2" - 8dC H6
0.1500	0.1950	0.4990[b]	0.7630[b]	0.240	0.330	8dC 2" - 9dC H6
0.0615	0.1490	0.2750	0.3590	0.280	0.330	8dC 2' - 8dC 1'
0.0000	0.0220	0.0940	0.1900	0.390	0.500	8dC 2' - 8dC H5
0.4030	0.5940	0.6910	0.6500	0.190	0.220	8dC 2' - 8dC H6
0.0235	0.0460	0.0820	0.1020	0.320	0.380	8dC 3' - 8dC 1'
0.0740	0.1150	0.3110	0.4000	0.270	0.310	8dC 3' - 8dC 2"
0.3020	0.3320	0.6140	0.5820	0.210	0.240	8dC 3' - 8dC 2'
0.0000	0.0400	0.0950	0.1750	0.330	0.420	8dC 3' - 8dC H6
0.0000	0.0000	0.0000	0.0240[b]			8dC 4' - 8dC H5
0.0185	0.0450	0.0630	0.0910	0.360	0.420	8dC H6 - 9dC H5
0.0000	0.0000	0.0000	0.0170			8dC H6 - 9dC H6
0.0130	0.0300	0.0580	0.1200	0.380	0.460	9dC 1' - 10dG H8
0.3320	0.5000	0.8180	1.1660	0.210	0.240	9dC 2" - 9dC 1'
0.0500	0.1950	0.4990[b]	0.7630[b]	0.260	0.390	9dC 2" - 9dC H6
0.0300	0.2070	0.2810[b]	0.3170[b]	0.250		9dC 2' - 9dC 1'
0.5820	0.8380	0.9410[b]	1.1600[b]	0.210	0.250	9dC 2' - 9dC H6
0.0455	0.0440	0.0600	0.0610	0.290	0.340	9dC 2' - 10dG H8
0.0160	0.0600	0.0890	0.1150	0.340	0.550	9dC 3' - 9dC 1'
0.0000	0.0000	0.0830	0.1550	0.300	0.450	9dC 3' - 9dC H6
0.0000	0.0230	0.0192[b]	0.0220[b]		0.450	9dC 3' - 10dG H8
0.0000	0.0000	0.0000	0.0120			9dC H6 - 10dG H8
0.0163	0.0360	0.0590	0.1450	0.390	0.430	10dG 1' - 10dG H8
0.4050	0.5320	0.9730	1.2520	0.200	0.240	10dG 2" - 10dG 1'
0.0450	0.2770	0.3200[b]	0.6190[b]	0.260	0.350	10dG 2' - 10dG 1'
0.2700	0.4460	0.6540	0.4650	0.210	0.230	10dG 2' - 10dG H8
0.0110	0.0405	0.0950	0.1440	0.370	0.460	10dG 3' - 10dG 1'
0.2030	0.3490	0.5030	0.6850	0.220	0.260	10dG 3' - 10dG 2'
0.0345	0.0860	0.1640	0.2450	0.300	0.360	10dG 3' - 10dG H8
0.0830	0.1590	0.2440		0.260	0.300	10dG 4' - 10dG 1'
0.1590	0.3070	0.3750	0.6760	0.240	0.280	10dG 4' - 10dG 3'
0.0000	0.0170	0.0220	0.0460	0.360	0.470	10dG 4' - 10dG H8
0.0665	0.0980	0.0980	0.3030	0.320	0.360	11dC 1' - 11dC H6
0.0000	0.0000	0.0150	0.0370	0.400	0.590	11dC 1' - 12dG 4'
0.0150	0.0220	0.0360	0.0670	0.400	0.440	11dC 1' - 12dG H8
0.2910	0.5530	0.7300	0.9940	0.210	0.240	11dC 2" - 11dC 1'
0.0505	0.1030	0.1600	0.3390	0.280	0.370	11dC 2" - 11dC H6
0.0510	0.1590	0.1950	0.4070	0.290	0.330	11dC 2' - 11dC 1'
0.0000	0.0000	0.0760	0.1530	0.310	0.440	11dC 2' - 11dC H5
0.3335	0.4480	0.6460	0.7640	0.210	0.240	11dC 2' - 11dC H6
0.0647	0.1090	0.1160	0.2100	0.260	0.320	11dC 2' - 12dG H8
0.0000	0.0460	0.0620	0.0980	0.310	0.370	11dC 3' - 11dC 1'
0.2290	0.3430	0.5530	0.6700	0.220	0.260	11dC 3' - 11dC 2'
0.0000	0.0000	0.0000	0.0190			11dC 3' - 11dC H5
0.0000	0.0410	0.0530	0.1510	0.310	0.460	11dC 3' - 11dC H6
0.0060	0.0000	0.0100	0.0340	0.400	0.500	11dC 4' - 11dC H6
0.0000	0.0000	0.0000	0.0320			11dC H6 - 12dG H8
0.0800	0.1420	0.2500	0.2710	0.290	0.350	12dG 1' - 12dG 2'
0.2700	0.5800	0.6290[b]	0.6800[b]	0.220	0.270	12dG 1' - 12dG 2"
0.3040	0.5320	0.7200	0.8870	0.210	0.230	13dG 2" - 13dG 1'
0.0130	0.0350	0.0620	0.1600	0.380	0.460	12dG 1' - 12dG H8
0.0000	0.0000	0.0000	0.0210	0.400	0.500	12dG 1' - 13dG 1'
0.0000	0.0000	0.0200	0.0330	0.380	0.500	12dG 1' - 13dG 4'
0.0460	0.1310	0.1705	0.2830	0.300	0.370	12dG 1' - 13dG H8
0.0280	0.0820	0.1210	0.2350	0.310	0.380	12dG 3' - 12dG 1'

0.0100	0.0380	0.0590	0.1660	0.380	0.480	12dG 3' - 12dG H8
0.0455	0.0960	0.1730	0.1460	0.280	0.330	12dG 4' - 12dG 1'
0.1550	0.3500	0.5700	0.6900	0.230	0.290	12dG 4' - 12dG 3'
0.0000	0.0120	0.0180	0.0410	0.390	0.500	12dG 4' - 12dG H8
0.0000	0.0000	0.0000	0.0140			12dG 4' - 13dG H8
0.0095	0.0190	0.0960	0.1640	0.400	0.490	13dG 1' - 13dG H8
0.0000	0.0000	0.0035[b]	0.0570	0.400	0.470	13dG 1' - 14dT 1'
0.0000	0.0000	0.0000	0.0190			13dG 1' - 14dT 3'
0.0275	0.0590	0.0820	0.1750	0.350	0.390	13dG 1' - 14dT H6
0.0000	0.0280	0.0900	0.1670	0.460	0.600	13dG 1' - 14dT HM
0.1500	0.2190	0.3190	0.3770	0.240	0.280	13dG 2" - 14dT H6
0.1100	0.2740	0.4200	0.6600	0.250	0.300	13dG 2' - 13dG 1'
0.2830	0.4580	0.5060	0.5460	0.210	0.240	13dG 2' - 13dG H8
0.0240	0.0650	0.1090	0.2050	0.340	0.420	13dG 3' - 13dG 1'
0.0255	0.0470	0.1230	0.2530	0.330	0.400	13dG 3' - 13dG H8
0.0610	0.1205	0.1770	0.1860	0.270	0.310	13dG 4' - 13dG 1'
0.0000	0.0140	0.0290	0.0470	0.370	0.460	13dG 4' - 13dG H8
0.0000	0.0100	0.0240	0.0510	0.460	0.540	13dG H8 - 14dT H6
0.0740	0.1690	0.2100	0.3090	0.310	0.400	13dG H8 - 14dT HM
0.0200	0.0565	0.1080	0.1740	0.370	0.430	14dT 1' - 14dT H6
0.0000	0.0000	0.0000	0.0270			14dT 1' - 15dG 3'
0.0000	0.0000	0.0160	0.0360	0.400	0.550	14dT 1' - 15dG 4'
0.0135	0.0390	0.0550	0.0960	0.389	0.450	14dT 1' - 15dG H8
0.2920	0.4610	0.7030	0.7840	0.210	0.240	14dT 2" - 14dT 1'
0.0765	0.1480	0.2230	0.4990	0.270	0.350	14dT 2" - 14dT H6
0.0880	0.1320	0.2030	0.1850	0.250	0.280	14dT 2" - 15dG H8
0.0935	0.1720	0.3610	0.5320	0.270	0.330	14dT 2' - 14dT 1'
0.4465	0.5510	0.5800	0.8870	0.200	0.230	14dT 2' - 14dT H6
0.0515	0.0760	0.1030	0.1150	0.280	0.330	14dT 2' - 15dG H8
0.0110	0.0420	0.0760	0.1890	0.350	0.420	14dT 3' - 14dT 1'
0.1730	0.3580	0.6150		0.230	0.280	14dT 3' - 14dT 2'
0.0000	0.0000	0.0260	0.0690	0.400	0.520	14dT 3' - 15dG H8
0.0195	0.0540	0.1240	0.1520	0.330	0.400	14dT H6 - 14dT 3'
0.0000	0.0145	0.0290	0.0360	0.410	0.490	14dT H6 - 15dG H8
0.0095	0.0520	0.0525	0.1410	0.370	0.460	15dG 1' - 15dG H8
0.0510	0.1340	0.1920	0.2580	0.300	0.370	15dG 1' - 16dA H8
0.2080	0.4170	0.4620[b]	0.7270[b]	0.220	0.300	15dG 2" - 15dG 1'
0.0820	0.1240	0.3180	0.3420	0.270	0.370	15dG 2" - 15dG H8
0.1190	0.1750	0.2920	0.3290	0.240	0.270	15dG 2" - 16dA H8
0.0780	0.2030	0.2610	0.4350	0.270	0.370	15dG 2' - 15dG 1'
0.3005	0.4390	0.6460	0.4990	0.205	0.230	15dG 2' - 15dG H8
0.0485	0.0790	0.1640	0.1990	0.290	0.340	15dG 2' - 16dA H8
0.0200	0.0580	0.1010	0.2190	0.360	0.400	15dG 3' - 15dG 1'
0.0103	0.0410	0.0920	0.1300	0.380	0.460	15dG 3' - 15dG H8
0.0520	0.1150	0.1610	0.2120	0.270	0.330	15dG 4' - 15dG 1'
0.0000	0.0000	0.0130	0.0420	0.450	0.630	15dG 4' - 15dG H8
0.0000	0.0000	0.0000	0.0120			15dG 4' - 16dA H8
0.0000	0.0000	0.0150	0.0450	0.460	0.780	15dG H8 - 16dA H8
0.0095	0.0420	0.0730	0.1390	0.390	0.490	16dA 1' - 16dA H8
0.0195	0.0370	0.1060	0.1520	0.370	0.420	16dA 1' - 17dT H6
0.0000	0.0210	0.0500	0.0780[b]	0.450	0.590	16dA 1' - 17dT HM
0.0730	0.1170	0.2500	0.3310	0.270	0.350	16dA 2" - 16dA H8
0.2220	0.2700	0.2170	0.3800	0.220	0.250	16dA 2" - 17dT H6
0.0185	0.0760	0.0790	0.2750	0.370	0.530	16dA 2" - 17dT HM
0.0860	0.2260	0.2700	0.3020	0.270	0.370	16dA 2' - 16dA 1'
0.2855	0.3620	0.5270	0.5170	0.200	0.240	16dA 2' - 16dA H8
0.0630	0.1360	0.2010	0.2650	0.260	0.360	16dA 4' - 16dA 1'
0.0090	0.0350	0.1120	0.1370	0.380	0.500	16dA 4' - 16dA 2"

0.0035	0.0190	0.0140	0.0480	0.380	0.520	16dA 4' - 16dA H8
0.0000	0.0000	0.0180	0.0420	0.400	0.630	16dA 4' - 17dT H6
0.0000	0.0000	0.0000	0.0200[b]			16dA H8 - 17dT 4'
0.0000	0.0125	0.0235	0.0550	0.450	0.540	16dA H8 - 17dT H6
0.0900	0.1450	0.2650	0.3960	0.320	0.410	16dA H8 - 17dT HM
0.0170	0.0310	0.0835	0.1220	0.370	0.420	17dT 1' - 17dT H6
0.0000	0.0220	0.0380	0.0640	0.410	0.450	17dT 1' - 18dA H8
0.2730	0.4020	0.5840	0.6970	0.210	0.250	17dT 2" - 17dT 1'
0.0475	0.1020	0.1260	0.2820	0.290	0.370	17dT 2" - 17dT H6
0.0557	0.0870	0.1050	0.1270	0.260	0.310	17dT 2" - 18dA H8
0.0470	0.1400	0.2200	0.3630	0.290	0.330	17dT 2' - 17dT 1'
0.3485	0.4800	0.6280	0.5620	0.200	0.240	17dT 2' - 17dT H6
0.0285	0.0600	0.0740	0.0760	0.310	0.360	17dT 2' - 18dA H8
0.0110	0.0680	0.0430	0.0680	0.320	0.500	17dT 3' - 17dT H6
0.0000	0.0000	0.0000	0.0100		0.450	17dT 3' - 18dA H8
0.0000	0.0000	0.0110	0.0290	0.530	0.640	17dT H6 - 18dA H8
0.0215	0.0410	0.0610	0.1340	0.370	0.410	18dA 1' - 18dA H8
0.0085	0.0510	0.0630	0.1230	0.350	0.480	18dA 1' - 19dG H8
0.3000	0.4840	0.7100	0.7340	0.210	0.240	18dA 2" - 18dA 1'
0.0515	0.1640	0.3140	0.4590	0.290	0.390	18dA 2" - 18dA H8
0.0710	0.1090	0.1840	0.2360	0.260	0.300	18dA 2" - 19dG H8
0.0870	0.2360	0.3280	0.4770	0.260	0.310	18dA 2' - 18dA 1'
0.3705	0.5130	0.6100	0.6010	0.200	0.220	18dA 2' - 18dA H8
0.0250	0.0270	0.0980	0.1740	0.310	0.410	18dA 3' - 18dA 1'
0.0108	0.0330	0.0700	0.1450	0.380	0.450	18dA 3' - 18dA H8
0.0000	0.0000	0.0290	0.0400	0.350	0.460	18dA 3' - 19dG H8
0.0555	0.1010	0.2420	0.3160	0.280	0.330	18dA 4' - 18dA 1'
0.0130	0.0200	0.0360	0.0560	0.360	0.430	18dA 4' - 18dA 2'
0.0000	0.0050	0.0150	0.0500	0.490	0.650	18dA 4' - 18dA H8
0.0000	0.0000	0.0030[b]	0.0310[b]			18dA H2 - 4dA 1'
0.0000	0.0130	0.0265[b]	0.0270[b]			18dA H2 - 18dA 1'
0.0000	0.0000	0.0115	0.0500			18dA H8 - 19dG H8
0.0330	0.0700	0.1390	0.1830	0.350	0.390	19dG 1' - 19dG H8
0.0000	0.0000	0.0240	0.0340	0.360	0.480	19dG 1' - 20dA 1'
0.0400	0.0730	0.1200	0.1610	0.330	0.370	19dG 1' - 20dA H8
0.2750	0.3830	0.5810	0.8490	0.220	0.250	19dG 2" - 19dG 1'
0.2710	0.3250	0.4980	0.5730	0.210	0.240	19dG 2' - 19dG H8
0.0000	0.0260	0.0690	0.1920	0.350	0.460	19dG 3' - 19dG 1'
0.2280	0.3610	0.6140		0.210	0.260	19dG 3' - 19dG 2'
0.0063	0.0460	0.0780	0.1250	0.340	0.470	19dG 3' - 19dG H8
0.0000	0.0000	0.0230	0.0420	0.400	0.550	19dG 3' - 20dA H8
0.0555	0.1310	0.2520	0.1460	0.270	0.320	19dG 4' - 19dG 1'
0.0160	0.0300	0.0510	0.1180	0.350	0.410	19dG 4' - 19dG 2"
0.0160	0.0280	0.0620	0.1130	0.360	0.420	19dG 4' - 19dG 2'
0.1300	0.2000	0.3260	0.7000	0.250	0.290	19dG 4' - 19dG 3'
0.0000	0.0000	0.0000	0.0480			19dG 4' - 19dG H8
0.0000	0.0000	0.0000	0.0310			19dG H8 - 20dA H8
0.0210	0.0150	0.0185	0.0340	0.380	0.460	20dA 1' - 20dA H2
0.0255	0.0435	0.0725	0.1410	0.370	0.400	20dA 1' - 20dA H8
0.3280	0.5370	0.8140	0.9290	0.210	0.240	20dA 2" - 20dA 1'
0.0530	0.1190	0.2060	0.3020	0.310	0.450	20dA 2" - 20dA H8
0.0890	0.1990	0.3020	0.5040	0.270	0.300	20dA 2' - 20dA 1'
0.0065	0.0190	0.0730	0.0600	0.400	0.550	20dA 2' - 20dA H2
0.3235	0.4360	0.7280	0.8630	0.210	0.230	20dA 2' - 20dA H8
0.0100	0.0150	0.0665	0.1440	0.370	0.470	20dA 3' - 20dA 1'
0.1185	0.1400	0.2710	0.3870	0.250	0.290	20dA 3' - 20dA 2"
0.1570	0.2820	0.4520	0.6020	0.240	0.270	20dA 3' - 20dA 2'
0.0185	0.0480	0.0970	0.1640	0.320	0.400	20dA 3' - 20dA H8

0.0685	0.1730	0.2440	0.4080	0.270	0.320	20dA 4' - 20dA 1'
0.0780	0.2100	0.2580	0.4880	0.260	0.300	20dA 4' - 20dA 3'
0.0100	0.0050	0.0300	0.0500	0.370	0.480	20dA 4' - 20dA H8

2. Overlapping cross-peaks

0.2650	0.4400	0.5800	0.8000	0.220	0.250	1dT 2" - 1dT 1'
						2dC 2" - 2dC 1'[d]
0.0500	0.1000	0.2300	0.4000	0.300	0.400	1dT 2' - 1dT 1'
						2dC 2' - 2dC 1'[d]
0.2490	0.4680	0.6560[b]	0.8740[b]	0.260		3dT 2" - 3dT H6
						2dC 2" - 3dT H6[d]
0.1250	0.2560	0.2900	0.6500	0.250	0.290	4dA 4' - 4dA 3'
						13dG 4' - 13dG 3'[d]
						16dA 4' - 16dA 3'[d]
0.0000	0.0270	0.0400[b]	0.1000[b]			4dA H2 - 5dT 1'
						4dA H2 - 18dA 1'[c]
0.0000	0.0000	0.0270	0.0850	0.430	0.550	5dT 1' - 6dC 1'
						17dT 1' - 18dC 1'[c]
						18dT 1' - 19dC 1'[c]
0.3050	0.5280	0.7140[b]	0.9960[b]			6dC H5 - 6dC H6
						3dT 1' - 3dT H6[c]
0.0190	0.0350	0.0210	0.0880	0.340	0.500	6dC 1' - 6dC 3'
						17dT 1' - 17dT 3'[d]
0.0100	0.0680	0.1180	0.2155	0.370	0.470	7dA 3' - 7dA 1'
						16dA 3' - 16dA 1'[d]
0.1500	0.3200	0.4410	0.6000	0.210	0.280	7dA 3' - 7dA 2"
						16dA 3' - 16dA 2"[d]
0.3200	0.5850	0.8050[b]	1.0310[b]			9dC H5 - 9dC H6
						9dC 1' - 9dC H6[c]
0.0810	0.1570	0.2800	0.4300	0.270	0.340	11dC 3' - 11dC 2"
						10dG 3' - 10dG 2"[d]
0.3570	0.7300	0.5950[b]	1.1850[b]	0.210	0.280	12dG 2' - 12dG H8[c]
						12dG 2" - 12dG H8
0.0106	0.0330	0.0895	0.1375	0.340	0.500	13dG 4' - 13dG 2'
						16dA 4' - 16dA 2'[d]
0.0475	0.1240	0.1425	0.2435	0.270	0.370	13dG 2' - 14dT H6
						16dA 2' - 17dT H6[d]
0.3200	0.4000	0.8300	0.8000	0.200	0.250	13dG 3' - 13dG 2'
						16dA 3' - 16dA 2'[d]
						18dA 3' - 18dA 2'[d]
0.1830	0.3830	0.5820	0.1260	0.250	0.400	13dG H8 - 12dG 2'
						13dG H8 - 12dG 2"[c]
						13dG H8 - 13dG 2"[c]
0.2630	0.3880	0.6410	0.8100			14dT H6 - 14dT HM
						17dT H6 - 17dT HM[d]
0.1450	0.3100	0.3400	0.7000	0.250	0.290	15dG 4' - 15dG 3'
						18dA 4' - 18dA 3'[d]
0.0295	0.0770	0.1040[b]	0.2450[b]	0.290	0.400	16dA H8 - 16dA 3'
						16dA H8 - 15dG 3'[c]
0.0120	0.0205	0.0405	0.0655	0.350	0.440	16dA 3' - 17dT H6
						13dG 3' - 14dT H6[d]
0.3400	0.5000	0.6950	0.8000	0.200	0.240	16dA 2" - 16dA 1'
						7dA 2" - 7dA 1'[d]
0.0900	0.2400	0.2900	0.5250	0.250	0.370	18dA 2' - 19dG H8
						19dG 2" - 19dG H8[c]
						13dG H8 - 13dG 2"[c]

3. Covalently bound proton pairs

0.1950	0.3330	0.4000	0.5120		1dT H6 - 1dT HM
0.3090	0.5150	0.6860	0.9270	[0.246]	2dC H6 - 2dC H5
0.2420	0.4190	0.6020	0.7980		3dT H6 - 3dT HM
0.9770	1.4230	1.5830	1.1100	[0.177]	4dA 2" - 4dA 2'
0.1830	0.3880	0.5850	0.8100		5dT H6 - 5dT HM
1.1940	1.8900	1.6630	1.6490	[0.177]	7dA 2" - 7dA 2'
0.2855	0.3990	0.7310	0.8970	[0.246]	8dC H6 - 8dC H5
1.0830	1.6700	2.3300	1.7000	[0.177]	10dG 2" - 10dG 2'
0.7610	1.2000	1.4790	1.4410	[0.177]	11dC 2" - 11dC 2'
0.3210	0.5550	0.6870	1.1060	[0.246]	11dC H6 - 11dC H5
1.0560	1.5050	1.6290	1.2470	[0.177]	16dA 2" - 16dA 2'
0.7120	1.0180	1.1290	0.9630	[0.177]	17dT 2" - 17dT 2'
1.1720	1.6000	1.9000	1.7000	[0.177]	18dA 2" - 18dA 2'
					3dG 2" - 13dG 2'[d]
0.9310	1.3600	1.5800	1.4800	[0.177]	1dT 2" - 1dT 2'
					2dC 2" - 2dC 2'[d]
					3dT 2" - 3dT 2'[d]
0.9100	1.2450	1.4700	1.3700	[0.177]	5dT 2" - 5dT 2'
					6dT 2" - 6dT 2'[d]
					8dT 2" - 8dT 2'[d]
					9dT 2" - 9dT 2'[d]
					14dT 2" - 14dT 2'[d]

B. NOE intensities and distances for exchangeable protons[e]

NOE intensities				Distance (nm.)		Proton pair
.1	.2	.3	.5 sec.			
1.25	2.55	5.6	6.55	0.350	0.410	3dT H3 - 19dG H1
1.05	2.35	3.4	3.8	0.350	0.410	3dT H3 - 17dT H3
0.9	2.35	4.8	5.8	0.350	0.410	15dG H1 - 5dT H3
1.3	2.95	4.2	6.15	0.350	0.410	15dG H1 - 14dT H3
1.3	2.45	3.85	6.55	0.360	0.410	13dG H1 - 14dT H3
1.2	3.55	3.95	5.85	0.350	0.410	13dG H1 - 12dG H1
5.3	11.9	17.5	26.7	0.280	0.320	3dT H3 - 18dA H2
1.4	1.9	2.0	5.7	0.380	0.460	3dT H3 - 4dA H2
1.8	3.25	4.0	7.45	0.320	0.440	15dG H1 - 7dA H2
						15dG H1 - 16dA H2[d]
8.0	13.7	21.6	30.4	0.270	0.310	14dT H3 - 7dA H2
7.8	20.0	31.7	45.2	0.250	0.290	5dT H3 - 16dA H2
6.6	11.5	15.6	24.1	0.280	0.320	17dT H3 - 4dA H2
1.4	3.0	2.8	9.9	0.360	0.440	17dT H3 - 18dA H2
0.7	1.2	3.8	6.1	0.360	0.440	13dG H1 - 7dA H2
0.9	2.4	6.4	11.1	0.320	0.400	19dG H1 - 20dA H2
0.9	2.4	1.7	5.3	0.380	0.460	19dG H1 - 18dA H2
5.8	9.7	12.0	14.3	0.230	0.270	15dG H1 - 6dC 4a
0.0	0.0	1.4	2.0	0.380	0.440	3dT H3 - 2dC 4a
0.0	0.7	1.0	1.7	0.370	0.430	14dT H3 - 6dC 4a
0.0	1.1	1.7	1.2	0.340	0.400	14dT H3 - 8dC 4a
1.1	2.3	2.5	3.9	0.320	0.400	5dT H3 - 6dC 4a
0.0	1.40	2.1	2.8	0.330	0.390	12dG H1 - 8dC 4a
5.0	9.0	10.2	13.8	0.230	0.270	12dG H1 - 9dC 4a

0.6	0.8	0.8	1.2	0.380	0.440	13dG H1 - 9dC 4a
3.2	6.8	9.5	13.5	0.240	0.280	13dG H1 - 8dC 4a
3.4	6.9	13.2	13.5	0.230	0.270	19dG H1 - 2dC 4a

C. Dihedral Angle restraints

Residue	ν_0		ν_1		ν_2		ν_3		ν_4	
all	-35	-15	-35	27	-35	-09	-13	25	-05	30
	α		β		γ		ε		ζ	
all[f]	-90	-30	135	-145	30	90	140	-60	150	-45
	δ									
1,11	30	90								

D. Major Groove distance restraints (nm.)

4dA C1' - 19dG C1'	> 0.70	5dT C1' - 19dG C1'	> 0.80
6dC C1' - 18dA C1'	> 0.80	7dA C1' - 17dT C1'	> 0.80
7dA C1' - 18dA C1'	> 1.10	8dC C1' - 16dA C1'	> 0.80

[a]The absence of an entry indicates that no distance estimate was made, or that the NOE intensity was not tabulated. NOE intensities and distances were not used for 5' and 5" protons because they were not assigned stereospecifically, and because the ratio of $J_{5'5"}$ to the difference in chemical shifts is often not small, leading to second order effects. [b]NOE noted in subsequent analysis of the data, but not used in structure refinement. [c]NOE and distances are included with the previous entry. [d]NOE and distances are approximately equal to the previous entry. [e]The one-dimensional difference NOEs are in percentage of the height of the irradiated imino peak. Imino-imino(i,j) NOE intensities are an average of the NOE observed on j while irradiating i, and the NOE observed on i, when irradiating j. [f]where applicable. For example, residue 1 has no 5' phosphate, and therefore, no corresponding α angle.

E. Coordinates (nm.) for d(T-C-T-A-T-C-A-C-C-G)•d(C-G-G-T-G-A-T-A-G-A)
404 atoms, L10 DNA, protons omitted

Atom	No.	x	y	z	Atom	No.	x	y	z
1dT O5*	1	0.094	0.436	1.554	11dC O5*	198	0.623	-0.105	-1.528
1dT C5*	2	0.196	0.480	1.644	11dC C5*	199	0.687	-0.002	-1.604
1dT C4*	3	0.332	0.419	1.611	11dC C4*	200	0.642	0.136	-1.559
1dT O4*	4	0.323	0.274	1.618	11dC O4*	201	0.497	0.143	-1.562
1dT C1*	5	0.369	0.219	1.493	11dC C1*	202	0.455	0.209	-1.442
1dT N1	6	0.290	0.097	1.459	11dC N1	203	0.316	0.166	-1.408
1dT C6	7	0.149	0.099	1.465	11dC C6	204	0.285	0.031	-1.387
1dT C2	8	0.354	-0.023	1.428	11dC C2	205	0.211	0.259	-1.412
1dT O2	9	0.475	-0.028	1.405	11dC O2	206	0.234	0.379	-1.417
1dT N3	10	0.278	-0.138	1.409	11dC N3	207	0.085	0.217	-1.399
1dT C4	11	0.139	-0.135	1.416	11dC C4	208	0.053	0.089	-1.380
1dT O4	12	0.077	-0.240	1.399	11dC N4	209	-0.075	0.052	-1.377
1dT C5	13	0.074	-0.015	1.443	11dC C5	210	0.153	-0.008	-1.373
1dT C5M	14	-0.078	-0.003	1.443	11dC C2*	211	0.556	0.170	-1.337
1dT C2*	15	0.347	0.327	1.391	11dC C3*	212	0.685	0.165	-1.416

1dT	C3*	16	0.377	0.453	1.470	11dC	O3*	213	0.757	0.290	-1.411
1dT	O3*	17	0.518	0.485	1.469	12dG	P	214	0.828	0.345	-1.277
2dC	P	18	0.594	0.540	1.337	12dG	O1P	215	0.855	0.231	-1.185
2dC	O1P	19	0.496	0.605	1.247	12dG	O2P	216	0.941	0.432	-1.316
2dC	O2P	20	0.713	0.616	1.380	12dG	O5*	217	0.710	0.434	-1.212
2dC	O5*	21	0.640	0.402	1.268	12dG	C5*	218	0.677	0.562	-1.268
2dC	C5*	22	0.746	0.323	1.324	12dG	C4*	219	0.564	0.628	-1.190
2dC	C4*	23	0.781	0.207	1.232	12dG	O4*	220	0.445	0.548	-1.200
2dC	O4*	24	0.666	0.121	1.219	12dG	C1*	221	0.385	0.546	-1.069
2dC	C1*	25	0.653	0.085	1.081	12dG	N9	222	0.297	0.427	-1.058
2dC	N1	26	0.510	0.053	1.057	12dG	C4	223	0.165	0.435	-1.057
2dC	C6	27	0.409	0.149	1.077	12dG	N3	224	0.082	0.540	-1.044
2dC	C2	28	0.471	-0.080	1.040	12dG	C2	225	-0.050	0.523	-1.043
2dC	O2	29	0.552	-0.166	1.005	12dG	N2	226	-0.128	0.628	-1.022
2dC	N3	30	0.342	-0.114	1.050	12dG	N1	227	-0.105	0.396	-1.056
2dC	C4	31	0.245	-0.025	1.073	12dG	C6	228	-0.021	0.285	-1.069
2dC	N4	32	0.120	-0.066	1.092	12dG	O6	229	-0.068	0.172	-1.074
2dC	C5	33	0.277	0.109	1.085	12dG	C5	230	0.116	0.306	-1.068
2dC	C2*	34	0.705	0.206	1.004	12dG	N7	231	0.221	0.224	-1.074
2dC	C3*	35	0.817	0.256	1.092	12dG	C8	232	0.330	0.299	-1.067
2dC	O3*	36	0.947	0.205	1.056	12dG	C2*	233	0.503	0.544	-0.974
3dT	P	37	1.019	0.237	0.915	12dG	C3*	234	0.597	0.641	-1.043
3dT	O1P	38	0.964	0.363	0.859	12dG	O3*	235	0.579	0.778	-1.001
3dT	O2P	39	1.165	0.224	0.933	13dG	P	236	0.605	0.831	-0.851
3dT	O5*	40	0.964	0.115	0.827	13dG	O1P	237	0.675	0.728	-0.772
3dT	C5*	41	1.001	-0.020	0.859	13dG	O2P	238	0.664	0.966	-0.860
3dT	C4*	42	0.926	-0.118	0.771	13dG	O5*	239	0.454	0.842	-0.797
3dT	O4*	43	0.784	-0.108	0.793	13dG	C5*	240	0.367	0.945	-0.847
3dT	C1*	44	0.724	-0.143	0.667	13dG	C4*	241	0.232	0.942	-0.778
3dT	N1	45	0.583	-0.094	0.663	13dG	O4*	242	0.165	0.818	-0.806
3dT	C6	46	0.552	0.040	0.690	13dG	C1*	243	0.100	0.775	-0.685
3dT	C2	47	0.479	-0.187	0.656	13dG	N9	244	0.090	0.627	-0.688
3dT	O2	48	0.497	-0.299	0.609	13dG	C4	245	-0.023	0.559	-0.691
3dT	N3	49	0.349	-0.147	0.684	13dG	N3	246	-0.151	0.595	-0.674
3dT	C4	50	0.319	-0.014	0.714	13dG	C2	247	-0.251	0.506	-0.685
3dT	O4	51	0.204	0.015	0.747	13dG	N2	248	-0.376	0.544	-0.662
3dT	C5	52	0.421	0.080	0.714	13dG	N1	249	-0.223	0.373	-0.714
3dT	C5M	53	0.394	0.230	0.728	13dG	C6	250	-0.091	0.332	-0.732
3dT	C2*	54	0.813	-0.079	0.563	13dG	O6	251	-0.064	0.218	-0.767
3dT	C3*	55	0.952	-0.090	0.624	13dG	C5	252	0.010	0.427	-0.719
3dT	O3*	56	1.029	-0.198	0.567	13dG	N7	253	0.142	0.419	-0.730
4dA	P	57	1.079	-0.200	0.414	13dG	C8	254	0.190	0.541	-0.711
4dA	O1P	58	1.073	-0.064	0.356	13dG	C2*	255	0.188	0.825	-0.573
4dA	O2P	59	1.208	-0.274	0.408	13dG	C3*	256	0.246	0.954	-0.627
4dA	O5*	60	0.964	-0.289	0.345	13dG	O3*	257	0.178	1.072	-0.581
4dA	C5*	61	0.950	-0.428	0.381	14dT	P	258	0.176	1.119	-0.426
4dA	C4*	62	0.824	-0.488	0.320	14dT	O1P	259	0.299	1.071	-0.358
4dA	O4*	63	0.710	-0.410	0.362	14dT	O2P	260	0.143	1.263	-0.421
4dA	C1*	64	0.623	-0.398	0.249	14dT	O5*	261	0.052	1.033	-0.372
4dA	N9	65	0.542	-0.275	0.266	14dT	C5*	262	-0.080	1.054	-0.423
4dA	C4	66	0.413	-0.274	0.297	14dT	C4*	263	-0.173	0.945	-0.373
4dA	N3	67	0.319	-0.370	0.302	14dT	O4*	264	-0.129	0.815	-0.417
4dA	C2	68	0.196	-0.339	0.339	14dT	C1*	265	-0.171	0.724	-0.314
4dA	N1	69	0.160	-0.216	0.372	14dT	N1	266	-0.093	0.597	-0.324
4dA	C6	70	0.249	-0.117	0.369	14dT	C6	267	0.048	0.597	-0.319
4dA	N6	71	0.216	0.005	0.410	14dT	C2	268	-0.159	0.479	-0.359
4dA	C5	72	0.379	-0.144	0.330	14dT	O2	269	-0.281	0.467	-0.348
4dA	N7	73	0.488	-0.069	0.317	14dT	N3	270	-0.085	0.366	-0.394

4dA	C8	74	0.586	-0.149	0.276	14dT	C4	271	0.055	0.368	-0.392
4dA	C2*	75	0.715	-0.394	0.129	14dT	O4	272	0.115	0.269	-0.433
4dA	C3*	76	0.828	-0.485	0.169	14dT	C5	273	0.121	0.483	-0.352
4dA	O3*	77	0.814	-0.620	0.119	14dT	C5M	274	0.273	0.489	-0.337
5dT	P	78	0.801	-0.660	-0.037	14dT	C2*	275	-0.147	0.799	-0.185
5dT	O1P	79	0.853	-0.550	-0.121	14dT	C3*	276	-0.175	0.943	-0.221
5dT	O2P	80	0.856	-0.796	-0.055	14dT	O3*	277	-0.302	0.990	-0.173
5dT	O5*	81	0.641	-0.664	-0.053	15dG	P	278	-0.334	1.011	-0.016
5dT	C5*	82	0.564	-0.759	0.024	15dG	O1P	279	-0.208	1.031	0.058
5dT	C4*	83	0.415	-0.734	0.009	15dG	O2P	280	-0.442	1.112	-0.003
5dT	O4*	84	0.384	-0.599	0.048	15dG	O5*	281	-0.392	0.866	0.022
5dT	C1*	85	0.282	-0.553	-0.043	15dG	C5*	282	-0.517	0.822	-0.033
5dT	N1	86	0.278	-0.405	-0.045	15dG	C4*	283	-0.546	0.678	0.006
5dT	C6	87	0.393	-0.328	-0.072	15dG	O4*	284	-0.446	0.590	-0.050
5dT	C2	88	0.163	-0.339	0.001	15dG	C1*	285	-0.424	0.486	0.047
5dT	O2	89	0.056	-0.398	0.016	15dG	N9	286	-0.292	0.424	0.023
5dT	N3	90	0.166	-0.201	0.023	15dG	C4	287	-0.275	0.301	-0.024
5dT	C4	91	0.281	-0.127	-0.002	15dG	N3	288	-0.360	0.198	-0.036
5dT	O4	92	0.283	-0.008	0.029	15dG	C2	289	-0.319	0.080	-0.084
5dT	C5	93	0.394	-0.191	-0.051	15dG	N2	290	-0.406	-0.020	-0.092
5dT	C5M	94	0.520	-0.116	-0.094	15dG	N1	291	-0.187	0.062	-0.124
5dT	C2*	95	0.319	-0.615	-0.176	15dG	C6	292	-0.097	0.168	-0.113
5dT	C3*	96	0.370	-0.751	-0.136	15dG	O6	293	0.019	0.155	-0.151
5dT	O3*	97	0.269	-0.854	-0.142	15dG	C5	294	-0.142	0.288	-0.061
6dC	P	98	0.196	-0.902	-0.277	15dG	N7	295	-0.081	0.403	-0.033
6dC	O1P	99	0.273	-0.855	-0.395	15dG	C8	296	-0.173	0.484	0.020
6dC	O2P	100	0.164	-1.046	-0.266	15dG	C2*	297	-0.434	0.557	0.180
6dC	O5*	101	0.060	-0.816	-0.270	15dG	C3*	298	-0.543	0.660	0.157
6dC	C5*	102	-0.038	-0.844	-0.168	15dG	O3*	299	-0.673	0.619	0.201
6dC	C4*	103	-0.152	-0.745	-0.173	16dA	P	300	-0.709	0.578	0.353
6dC	O4*	104	-0.099	-0.612	-0.159	16dA	O1P	301	-0.608	0.635	0.446
6dC	C1*	105	-0.177	-0.527	-0.245	16dA	O2P	302	-0.852	0.605	0.377
6dC	N1	106	-0.097	-0.406	-0.277	16dA	O5*	303	-0.685	0.419	0.346
6dC	C6	107	0.027	-0.415	-0.343	16dA	C5*	304	-0.768	0.338	0.261
6dC	C2	108	-0.134	-0.282	-0.223	16dA	C4*	305	-0.727	0.192	0.269
6dC	O2	109	-0.247	-0.265	-0.176	16dA	O4*	306	-0.592	0.177	0.223
6dC	N3	110	-0.052	-0.177	-0.233	16dA	C1*	307	-0.530	0.079	0.309
6dC	C4	111	0.066	-0.184	-0.294	16dA	N9	308	-0.383	0.105	0.310
6dC	N4	112	0.146	-0.078	-0.295	16dA	C4	309	-0.291	0.024	0.260
6dC	C5	113	0.108	-0.303	-0.352	16dA	N3	310	-0.294	-0.103	0.217
6dC	C2*	114	-0.207	-0.614	-0.366	16dA	C2	311	-0.184	-0.158	0.169
6dC	C3*	115	-0.225	-0.752	-0.306	16dA	N1	312	-0.069	-0.092	0.160
6dC	O3*	116	-0.364	-0.784	-0.286	16dA	C6	313	-0.060	0.035	0.200
7dA	P	117	-0.468	-0.809	-0.407	16dA	N6	314	0.051	0.104	0.182
7dA	O1P	118	-0.394	-0.844	-0.530	16dA	C5	315	-0.173	0.096	0.253
7dA	O2P	119	-0.574	-0.900	-0.359	16dA	N7	316	-0.196	0.217	0.303
7dA	O5*	120	-0.529	-0.661	-0.426	16dA	C8	317	-0.324	0.221	0.338
7dA	C5*	121	-0.620	-0.607	-0.329	16dA	C2*	318	-0.593	0.100	0.444
7dA	C4*	122	-0.656	-0.463	-0.361	16dA	C3*	319	-0.734	0.143	0.413
7dA	O4*	123	-0.538	-0.380	-0.356	16dA	O3*	320	-0.832	0.039	0.424
7dA	C1*	124	-0.555	-0.280	-0.458	17dT	P	321	-0.869	-0.033	0.563
7dA	N9	125	-0.423	-0.225	-0.496	17dT	O1P	322	-0.833	0.055	0.676
7dA	C4	126	-0.383	-0.101	-0.472	17dT	O2P	323	-1.008	-0.085	0.554
7dA	N3	127	-0.446	0.009	-0.430	17dT	O5*	324	-0.766	-0.156	0.560
7dA	C2	128	-0.380	0.123	-0.419	17dT	C5*	325	-0.774	-0.256	0.458
7dA	N1	129	-0.251	0.133	-0.448	17dT	C4*	326	-0.653	-0.347	0.464
7dA	C6	130	-0.182	0.027	-0.489	17dT	O4*	327	-0.533	-0.271	0.446
7dA	N6	131	-0.052	0.036	-0.514	17dT	C1*	328	-0.432	-0.338	0.523

7dA	C5	132	-0.248	-0.094	-0.503	17dT	N1	329	-0.321	-0.244	0.555
7dA	N7	133	-0.210	-0.214	-0.547	17dT	C6	330	-0.346	-0.120	0.616
7dA	C8	134	-0.318	-0.293	-0.544	17dT	C2	331	-0.193	-0.269	0.505
7dA	C2*	135	-0.623	-0.353	-0.571	17dT	O2	332	-0.160	-0.381	0.465
7dA	C3*	136	-0.716	-0.449	-0.500	17dT	N3	333	-0.095	-0.170	0.511
7dA	O3*	137	-0.850	-0.397	-0.489	17dT	C4	334	-0.122	-0.046	0.569
8dC	P	138	-0.947	-0.366	-0.615	17dT	O4	335	-0.036	0.042	0.563
8dC	O1P	139	-0.891	-0.427	-0.738	17dT	C5	336	-0.248	-0.022	0.622
8dC	O2P	140	-1.086	-0.398	-0.577	17dT	C5M	337	-0.282	0.106	0.699
8dC	O5*	141	-0.930	-0.206	-0.628	17dT	C2*	338	-0.505	-0.386	0.647
8dC	C5*	142	-0.983	-0.120	-0.526	17dT	C3*	339	-0.645	-0.416	0.599
8dC	C4*	143	-0.950	0.026	-0.554	17dT	O3*	340	-0.673	-0.556	0.585
8dC	O4*	144	-0.807	0.041	-0.566	18dA	P	341	-0.685	-0.658	0.710
8dC	C1*	145	-0.781	0.122	-0.682	18dA	O1P	342	-0.691	-0.582	0.837
8dC	N1	146	-0.644	0.090	-0.732	18dA	O2P	343	-0.792	-0.756	0.680
8dC	C6	147	-0.613	-0.035	-0.786	18dA	O5*	344	-0.542	-0.731	0.702
8dC	C2	148	-0.540	0.183	-0.717	18dA	C5*	345	-0.508	-0.808	0.586
8dC	O2	149	-0.562	0.296	-0.671	18dA	C4*	346	-0.360	-0.844	0.585
8dC	N3	150	-0.415	0.151	-0.751	18dA	O4*	347	-0.285	-0.721	0.600
8dC	C4	151	-0.384	0.031	-0.800	18dA	C1*	348	-0.183	-0.744	0.699
8dC	N4	152	-0.257	0.001	-0.827	18dA	N9	349	-0.156	-0.614	0.766
8dC	C5	153	-0.482	-0.065	-0.819	18dA	C4	350	-0.050	-0.538	0.740
8dC	C2*	154	-0.893	0.090	-0.778	18dA	N3	351	0.064	-0.560	0.673
8dC	C3*	155	-1.011	0.075	-0.685	18dA	C2	352	0.152	-0.463	0.662
8dC	O3*	156	-1.081	0.199	-0.662	18dA	N1	353	0.134	-0.343	0.713
9dC	P	157	-1.150	0.287	-0.779	18dA	C6	354	0.022	-0.315	0.781
9dC	O1P	158	-1.183	0.200	-0.894	18dA	N6	355	0.003	-0.193	0.830
9dC	O2P	159	-1.256	0.370	-0.719	18dA	C5	356	-0.073	-0.414	0.796
9dC	O5*	160	-1.026	0.382	-0.821	18dA	N7	357	-0.191	-0.418	0.858
9dC	C5*	161	-0.972	0.475	-0.727	18dA	C8	358	-0.240	-0.541	0.840
9dC	C4*	162	-0.851	0.547	-0.785	18dA	C2*	359	-0.240	-0.849	0.792
9dC	O4*	163	-0.747	0.452	-0.818	18dA	C3*	360	-0.324	-0.935	0.701
9dC	C1*	164	-0.681	0.508	-0.933	18dA	O3*	361	-0.255	-1.051	0.651
9dC	N1	165	-0.601	0.404	-1.002	19dG	P	362	-0.204	-1.172	0.746
9dC	C6	166	-0.659	0.286	-1.050	19dG	O1P	363	-0.281	-1.172	0.873
9dC	C2	167	-0.461	0.416	-1.004	19dG	O2P	364	-0.198	-1.296	0.666
9dC	O2	168	-0.407	0.524	-0.982	19dG	O5*	365	-0.054	-1.123	0.780
9dC	N3	169	-0.386	0.314	-1.047	19dG	C5*	366	0.045	-1.118	0.676
9dC	C4	170	-0.439	0.200	-1.092	19dG	C4*	367	0.172	-1.050	0.724
9dC	N4	171	-0.359	0.099	-1.125	19dG	O4*	368	0.141	-0.915	0.765
9dC	C5	172	-0.577	0.183	-1.095	19dG	C1*	369	0.222	-0.889	0.881
9dC	C2*	173	-0.792	0.567	-1.016	19dG	N9	370	0.164	-0.776	0.957
9dC	C3*	174	-0.886	0.623	-0.911	19dG	C4	371	0.219	-0.656	0.963
9dC	O3*	175	-0.866	0.764	-0.890	19dG	N3	372	0.336	-0.510	0.916
10dG	P	176	-0.912	0.876	-0.996	19dG	C2	373	0.374	-0.484	0.938
10dG	O1P	177	-1.020	0.821	-1.082	19dG	N2	374	0.493	-0.444	0.895
10dG	O2P	178	-0.936	1.002	-0.924	19dG	N1	375	0.292	-0.397	1.009
10dG	O5*	179	-0.779	0.889	-1.086	19dG	C6	376	0.169	-0.442	1.057
10dG	C5*	180	-0.668	0.970	-1.045	19dG	O6	377	0.095	-0.366	1.120
10dG	C4*	181	-0.556	0.962	-1.147	19dG	C5	378	0.133	-0.574	1.034
10dG	O4*	182	-0.498	0.830	-1.148	19dG	N7	379	0.028	-0.647	1.069
10dG	C1*	183	-0.443	0.814	-1.280	19dG	C8	380	0.048	-0.770	1.023
10dG	N9	184	-0.436	0.669	-1.310	19dG	C2*	381	0.224	-1.021	0.956
10dG	C4	185	-0.324	0.602	-1.335	19dG	C3*	382	0.232	-1.122	0.843
10dG	N3	186	-0.197	0.642	-1.348	19dG	O3*	383	0.367	-1.161	0.813
10dG	C2	187	-0.100	0.554	-1.374	20dA	P	384	0.462	-1.244	0.914
10dG	N2	188	0.026	0.596	-1.385	20dA	O1P	385	0.382	-1.297	1.027
10dG	N1	189	-0.129	0.418	-1.390	20dA	O2P	386	0.545	-1.338	0.835

10dG	C6	190	-0.261	0.375	-1.377	20dA	O5*	387	0.555	-1.126	0.972
10dG	O6	191	-0.290	0.257	-1.397	20dA	C5*	388	0.655	-1.064	0.888
10dG	C5	192	-0.360	0.468	-1.349	20dA	C4*	389	0.708	-0.938	0.953
10dG	N7	193	-0.491	0.459	-1.333	20dA	O4*	390	0.600	-0.848	0.981
10dG	C8	194	-0.537	0.582	-1.311	20dA	C1*	391	0.638	-0.774	1.098
10dG	C2*	195	-0.535	0.887	-1.375	20dA	N9	392	0.516	-0.735	1.172
10dG	C3*	196	-0.604	0.990	-1.288	20dA	C4	393	0.491	-0.615	1.223
10dG	O3*	197	-0.570	1.123	-1.330	20dA	N3	394	0.560	-0.501	1.226
						20dA	C2	395	0.510	-0.397	1.290
						20dA	N1	396	0.393	-0.400	1.352
						20dA	C6	397	0.319	-0.511	1.352
						20dA	N6	398	0.200	-0.513	1.410
						20dA	C5	399	0.368	-0.624	1.287
						20dA	N7	400	0.323	-0.748	1.273
						20dA	C8	401	0.415	-0.816	1.205
						20dA	C2*	402	0.719	-0.869	1.184
						20dA	C3*	403	0.776	-0.969	1.085
						20dA	O3*	404	0.918	-0.950	1.076